# Workshop korpus českého znakového jazyka (4)

Jak zatím pokračují práce v Elanu?

* Žofka:
  + u druhého videa se nepodařilo propojit vrstvy
  + Elan viděla poprvé
  + problém se zpomalením videa
* Honza Š.:
  + problém neměl, přečetl si manuál, má zkušenosti z jiných střihačských programů
* Honza C.:
  + nepovedlo se nastavit to, aby fungoval vokabulář/controller vocabulary (CV)
  + není si jistý, jestli se mu podařilo uložit, to co zkoušel

Rekapitulace problematických míst:

* jak udělat, aby se v nové vrstvě zobrazovaly předdefinované hodnoty CV
  + **add tier** -> **tier name** - přidat novou vrstvu
  + změna vlastnosti vrstvi (**Change tier atrributes**) – **Tier type**: (tam nastavit „slovní druh“ (CV))
* 3 kroky:
  + CV -> TYPE -> TYPE -> TIER
    - První vytvořit CV,
    - vytvoří **type** a s tím se propojí ten **CV**
    - po té vytvoříme vrstvu (tier) a s tou propojíme ten **typ**
  + PŘ: Tvary ruky
    - **Edit** -> **Edit CV** -> zadat jednotlivé hodnoty, třeba čísly
    - **Type** -> **Type name** (+ **USE CV** – ten co už máme vytvořený)
    - **TIER** – „pravá ruka“ může se nastavit jméno účastníka a u „**TIER TYPE**“ – nastavíme handshape a přidat
* Jak přenést hranice znaku přes vrstvy (stejné hranice na různých vrstvách)? Existují dva způsoby:
  + jednodušší × více mechanice
    - kliknout na anotaci – zmodrá a označí se cely sloupec
    - v požadované vrstvě pravým tlačítkem –> **new annotation here** a ona se vytvoří
  + použití stereotypu
    - vlastnost, která se nastavuje u „TYPE“
    - základní možnost:
      * none: žádný (nová anotace může být kdekoliv a není vázána na jiné vrstvy)
      * time subdivision: vázána na jinou vrstvu, ale může být ještě dále dělena
      * symbolic associations: že se váže na již existující anotaci
  + musí se nastavit hierarchie mezi vrstvami
    - **change parent of tier**
      * to první, co označím, bude označeno jako podřazená vrstva -> (**next**)
      * -> nadřazená vrstva
      * nová vrstva z podřazené bude automaticky kopírovat hranice z nadřazené
* zpomalení: **rate**: nastavit menší hodnotu
* přehrání pouze vybrané části:
  + označit **play selection** (play+S) -> přehraje jen vybranou část
  + když **loop mode:** přehrávání ve smyčce
* vybrání více vrstev najednou – ručně myší (možná lze nějakou klávesovou zkratkou)

Příklad anotace z korpusu nizozemského znakového jazyka

* elicitační nahrávka převyprávění příběhu
* pozn.: transkripce jedné minuty mluvené produkce bez obrazu trvá cca 8 minut (u znakového jazyka mnohonásobně delší)
* struktura anotace
  + vrstva překladu (do aj, do nizozemštiny)
  + vrstva transkripce: glosy pro jednotlivé znaky (jak pro pravou, tak pro levou ruku)
    - těmto dvěma vrstvám jsou podřazeny dvě další vrstvy (umístění a směr pravé/levé ruky)
  + nezávislé překlady glos do angličtiny
    - u obouručních znaků bude ta glosa stejná (pro levou a pravou ruku ve stejné chvíli)
  + pohyby úst
  + obočí
  + otevřené/zavřené oči
  + směr pohledu
  + pohyb úst
    - vyskytují se dvakrát
    - spíše dva různé anotační systémy pro mluvní komponenty
    - v podstatě totožné, jen se asi liší použitou terminologií
    - pravděpodobně anotováno dvěma různými anotátory
  + anotace mimiky
  + pohybů hlavy
  + rolí (střídání rolí), změna perspektivy
  + komentáře
  + pozn.: nemáme vrstvu pro tvary ruky –> takže se ještě nejedná o finální verzi anotace
* jak je to ve finálních vrstvách s CV?
  + např. u vrstvy otevřené/zavřené očí:
    - použité značky × nemusí být snadno rozpoznatelné, co která značka znamená
    - w – wide – rozšířené oči
    - b – blink – mrkání
    - s – přimhouřené oči
  + má to nějakou distinktivní vlastnost?
    - když třeba oči do široka: význam nějakého zdůraznění
    - zúžené oči: odkaz na něco o čem už jsme mluvili
    - mrkání
      * jak rozeznat, kdy se jedná o bezděčnou činnost a kdy o jazykový jev
      * ve specifických případech mrknutí využito -> překvapení
      * buď nutné zajistit značky pro dva druhy mrknutí (jazykové × fyziologické), nebo k tomu fyziologickému přistupovat jako k bezděčnému pohybu (např. jako škrábání rukou na hlavě) a tak jej do anotace nezačleňovat vůbec
  + role:
    - Změny perspektivy např. já jsem přímo pes × popisuji okolí, kde se pes nachází, nebo situace
    - z jakého pohledu uchopím tu situaci: z pohledu mluvčího, nebo anotátora
      * tady asi bude docházet k simultaneitě rolí, které se budou překrývat
      * Př: člověk + pes: situace, kdy se dobře rozpozná kdo je kdo, ale jak postupovat v případě, kdy budou spolu reagovat dvě osoby, nebo dvě zvířata?
      * možná z vyprávění kontextu/celku (proto jeden anotátor anotuje jeden ucelený úsek)
      * anotaci by měli provádět, alespoň tři anotátoři, z nichž by alespoň jeden měl být rodilí mluvčí znakového jazyka, v případě neshody by měly probíhat vzájemné konzultace
  + otázka: Nebylo by vhodně, aby se anotace rozsegmentovala? Někdo by se zaměřil třeba pouze na tvary rukou, někdo na mluvní komponenty a pak by se jednotlivé části spojily?
  + × na druhou stranu, když jeden anotátor bude vyplňovat vrstvy napříč, může vidět, co spolu souvisí a lépe chápe kontext

Hierarchie mezi vrstvami a stereotypy

* stereotyp se netýká hierarchie, ale způsobu jakým se vytvářejí vrstvy
* vrstvy mohou být diskrétní
* v transkripci, kde nás nezajímajíc přesné hranice znaků, můžeme nastavit, že jednotlivé anotace na sebe budou bezprostředně navazovat,
  + pozn.: lze nastavit možnost vkládání dalších anotací přímo dovnitř již vytvořených anotací
* buď to
  + ve stereotypu –> lze nastavit dělitelnost jednotlivých anotací
  + stereotyp je nutné přiřadit k typu, v žádném jiném kroku to udělat nezle
  + když už máme vrstvu, nemůže v ní zpětně měnit stereotyp
* hierarchie:
  + nastavuje se pomocí nastavení nadřazené vrstvy (parent × child)
    - hierarchické vztahy je nutné nastavit, ještě předtím, než jsou ve vrstvě vytvořeny nějaké anotace, pak už se vztahy měnit nedají
    - nemůžeme je hierarchicky propojit
  + Př.: horní vrstva např. pohyb pravé ruky (včetně tranzitních pohybů) v podřazené vrstvě by po té byl celý pohyb rozfázován do jednotlivých částí

Elan

* může fungovat jako nástroj pro korpusové vyhledávání: záložka **vyhledávání**, nebo klávesová zkratka **ctrl+f**
* pozn.: lze nastavit konkrétní úsek (**distance of – to**)
  + lze vyhledávat, jestli se v úseku jev vyskytuje, nebo nevyskytuje (např. najdi úseky, kde se tvar ruky nevyskytuje v pěti po sobě následujících znacích)
  + využití intervalu je praktické zejména máme-li např. hodinovou
  + připomíná to hledání v korpusech
  + první kolonka: omezení vyhledávání je spousta (např. hledej v dané vrstvě)
  + první kolonka: co ano × ne
    - spuštění vyhledávacího procesu „**šipka pro přehrávání**“
    - vyskočí všechny nalezené výsledky
  + vyhledávací nástroj umí pracovat i s tzv. regulérními výrazy („**regular expressions**“)
    - tj. dotazovací jazyk pro vyhledávání v korpusu (v podstatě stejně jako v ČNK)
      * [] \| . ?

×

* export
  + vyhledávání v Elanu je vhodné spíše pro rychlé ověřování
  + výsledky lze vygenerovat do dokumentu, který obsahuje data
* export – **file** -> **export as**
  + nejdůležitější prostý textový soubor (formát txt), nebo csv
  + může se nastavit, které konkrétní informace chceme exportovat (anotátory, participanty, jednotlivé vrstvy, lze to omezit na určitý úsek anotace)
  + např vygenerování v tabulkovém formátu (vhodné pro další zpracování)
* export do html
  + pomáhá zpracovat vrstvy, elegantní přehled anotací
  + dá se otevřít v prohlížečích a následně prohledávat

Kromě technické části (přímé práce s Elanem), je potřeba vyřešit spoustu záležitostí týkajících se jednotlivých vrstev

Slovní druhy

* Ty co jsme si určily minule jsou nedostačující, o které je potřeba CV rozšířit? Např. o:
* Předložka
* Specifický znak
* Klasifikátor:
  + otázka – jestli použití klasifikátoru není součástí slovesného vyjádření/slovesa × kdy stojí samostatně?
  + pozn.: blending: neumíme odlišit, kde je verbální a substantivní část
* typ částic – ano/ne
  + v češtině se pojímají jako kontaktové částice/kontaktové citoslovce
  + diskursivní markery
  + kontaktová slova
* př. vyprávět: bude sloveso × nebo jako příběh
  + z kontextu
  + LYŽUJI RÁD × LYŽOVÁNÍ MÁM RÁD
  + jak to bude z deverbativy?
  + nutné nastavit interní pravidla
  + bude nás zajímat verbální jádro -> co bude mít verbální jádro, budeme pojímat jako verbum
    - když uděláme – může se objevit problém, že z frekvenčního hlediska znakový jazyk používá obrovské množství sloves –> zkreslil by se tím náhled na znakový jazyk
    - můžeme udělat vlastní kategorii
    - pozn.: u angličtiny také některá slova vypadají stejně × díky pevnému slovosledu lze snadněji rozpoznat slovní druh
    - úkol do příště popřemýšlet o kategoriích
  + dobré by bylo vydělit pomocná slovesa

Jaké slovní druhy budeme vyčleňovat:

* pomocná slovesa -fungují jinak než běžná slovesa (BUDE)
* konjunkce (PROTO, DŮVOD)
* částice (ASI)
* ANO, NE
* specifické znaky
* klasifikátory? (asi vyčlenit)
* nepřímá pojmenování- zdrojem NP mohou být různá slovní druhy, navíc význam se posouvá (asi vyčlenit)
* příklad: manuální produkce slovesa+mimika vyjadřuje další informace
  + Má se to popsat jako příslovce ve vrstvě pro slovní druhy?
  + Spíš ne, zařadíme to do vrstvy pro mimiku.
  + Jde o blending – hranice slovních druhů jsou neostré, s tím se setkáme často
* příklad z videa: ʺnasadit si klobouk ʺ Máme anotovat jako KLOBOUK, nebo NASADIT?
  + Návrh: NASADIT SI NA HLAVU.
* **ÚKOL**: PREMÝŠLET O SLOVNĚDRUHOVÝCH KATEGORIÍCH, ABYCHOM NA PŘÍŠTÍM SEMINÁŘI MOHLI ROZHODNOUT, JAK JE BUDEME POUŽÍVAT

Příklady z videa:

* simultaneita: každá ruka produkuje jiný znak (řídit a telefonovat) -> můžeme mít pro každou ruku samostatnou vrstvu
* gesto nebo znak? (dotyk na tváři- naznačující přemýšlení)
  + něco jako hesitační zvuky v mluveném jazyce – určeno jako gesto – můžeme mít samostatnou vrstvu pro gesta
* různá gesta:
  + Sprásknutí rukou (možná probíhá proces gramatikalizace)
  + Mnutí prstů
  + Drbání, posunování brýlí (gesto typické pro určitou osobu)

Praktická část hodiny:

* anotace slovních druhů
  + - přidat vrstvu pro pravou ruku a vrstvu pro levou ruku (Vrstva –> Přidat novou vrstvu)
* glosy
  + - přidat vrstvu pro pravou ruku a vrstvu pro levou ruku
* Vrstvy je možné přesunovat, například tak, aby vrstvy pro pravou ruku byly pod sebou
* Kliknutí na první sloupec pravým tlačítkem, vybíráme, jaké vrstvy chceme viditelné

Další možnosti nastavení:

* Provázanost vrstev- Bude v Moodle
* Stereotyp – určuje, jak jsou anotace ve vrstvě mezi sebou navázané
  + (Typ-Přidat nový typ vrstvy- v tabulce vybrat Stereotyp)
    - 1. nespecifikovaný
    - 2. anotace je vnitřně rozdělená na menší části
    - 3. anotace na sebe těsně navazují (přechodová část mezi znaky je zahrnuta)
    - 4. anotace jsou oddělené
    - 5 anotace jsou oddělené+lze mezi ně vložit další anotaci
* Pokud vybereme stereotyp, vyplníme nějakou anotaci a rozhodneme se, že chceme stereotyp změnit, už to není možné. To platí i pro nastavení provázanosti vrstev.
* pozn.: kontrolovaný slovník (CV) je možné měnit (Editovat- Editovat kontrolované slovníky)

Opakování: Nastavení kontrolovaného slovníku (probíhá ve třech krocích):

* **1) Editovat**: **Editovat kontrolované slovníky** –> v tabulce vyplnit Jméno např. slovní druhy –> v kolonce **Zadat hodnotu** napsat např. substantivum - kliknout **přidat** pod kolonkou Zadat hodnotu- vyplnit další hodnotu např. adjektivum atd.
  + nakonec kliknout **Přidat** napravo od kolonky pro Jméno- vytvoří se nový kontrolovaný slovník
* 2) **Typ** -> **Přidat nový typ** vrstvy- v tabulce vyplnit Jméno typu např. slovní druhy -> v kolonce **Použít kontrolované slovníky** vybrat Slovní druhy- kliknout **Přidat**-vytvoří se nový typ
* 3) **Vrstva** -> **Přidat novou vrstvu** -> v tabulce vyplnit Jméno vrstvy -> v kolonce **Typ vrstvy** vybrat Slovní druhy -> kliknout **Přidat**- vytvoří se nová vrstva

* Na webové stránce Elanu (<https://tla.mpi.nl/tools/tla-tools/elan/>) je manuál a také vzorové anotované video (+bude ke stažení také na Moodle)
* **ÚKOL**: POKRAČOVAT V ANOTACI, VČETNĚ NOVĚ VYTVOŘENÝCH VRSTEV PRO PRAVOU A LEVOU RUKU. KOLEM DESETI ZNAKŮ (kolik stihneme ☺ )