

Demonstrator scenario

Om te demonstreren hoe de FESLI-applicatie kan worden gebruikt, wordt hieronder gerefereerd aan een aantal zoekacties en onderzoeksvragen die in het FESLI-voorstel zijn genoemd en zoals die in de loop van het FESLI-proces zijn besproken. Ze zijn representatief voor vragen die binnen het BiSLI-onderzoek werden geformuleerd.

Hierbij worden resultaten getoond van zoekacties in de webapplicatie (<http://yago.meertens.knaw.nl/apache/Fesli/>) waarbij queries worden gebruikt die passen binnen de Corpus Query Language (<http://www.fi.muni.cz/~thomas/corpora/CQL/>). Een uitleg hiervan is te vinden in de *User Manual* (Deliverable 10) van Marc Kemps-Snijders, in de Project Area.

Belangrijk is daarbij dat in de applicatie een zoekopdracht kan worden uitgevoerd op drie niveaus: *pos* (part of speech), *lemma*, *word* (woord).

Hieronder worden enkele van de vragen (desiderata) uit (onder andere) het voorstel aangehaald en daarbij wordt de zoekroute exemplarisch weergegeven.

- (1) *We zouden uit de diverse corpora per kind, per stadium de gevallen van gebruik van het onzijdige lidwoord het (het glas) willen filteren en kwantificeren.*
- Smalle (item-specifieke) route: zoekopdracht ["het"] ["glas"]. Deze opdracht geeft 25 hits. De resultaten worden aangegeven per kind en ook verdeeld naar groep en naar al of niet taalgestoord. De hier gekozen route genereert alleen uitingen uit de elicitatietask, waarvan we immers weten dat het woord glas erin is bevraagd. Een ander onderdeel van de vraag is, geparafraseerd: hoe vinden we de data per stadium? Weliswaar zijn de BiSLI-data niet longitudinaal, maar de kinderen verschillen in leeftijd en een dergelijke vraag kan met de applicatie gesteld worden. Om deze vraag te beantwoorden kunnen waarden voor een leeftijdsrange worden gekozen in het uitklapmenu van het zoekscherm.
- Brede route: zoekopdracht "het" [pos='n.*onz.*'] (pos betekent: part of speech). Deze opdracht geeft 1018 hits. Tussen vierkante haken wordt aangegeven dat 'het' gevolgd wordt door een onzijdig naamwoord. Wanneer 'het' gebruikt wordt met een zijdig naamwoord (incorrect dus) vinden we 111 hits. Wanneer we in het uitklapmenu de optie 'Wel taalgestoord' aanvinken, zien we dat meeste fouten door deze groep worden gemaakt (73 hits). In deze uitkomsten zijn de data uit alle gebruikte taken (elicitering en narratie) betrokken. Het uitklapmenu maakt het ook mogelijk de taken afzonderlijk te beschouwen. Niettemin is het goed om te vermelden dat de FESLI-applicatie expliciet mede ontwikkeld is om verschillende soorten data (taken) aan eenzelfde zoekopdracht te onderwerpen. Bij de 38 foute combinaties kunnen we zichtbaar maken dat (anders dan bij de taalgestoorde groep, waar de fout breed voorkomt) de fouten bij de niet-taalgestoorde groep op het conto komen van vier individuele kinderen. Ook is het bij de taalgestoorde groep mogelijk te onderscheiden (via de knop voor Taal) hoeveel van de fouten worden gemaakt door respectievelijk de monolinguale en bilinguale kinderen.

- (2) *We zouden de gegevens van de adjectieven graag willen vergelijken met die van de lidwoorden en demonstratieven.*

Zoekopdrachten kunnen worden gegeven waarbij congruentie (adjectief-nomen) en *gender assignment* (bij lidwoorden en demonstrativa) worden gemeten. Voor de lidwoorden is de route hierboven weergegeven: de zijdigheid van een naamwoord kan worden vergeleken met het erbij gebruikte (bepaalde) lidwoord. Voor aanwijzende voornaamwoorden kan eenzelfde exercitie worden uitgevoerd, waarbij het voornaamwoord het lidwoord vervangt in de zoekopdracht. Zo worden naast zinsdelen als 'het bal' ook uitingen met 'dat bal' gevonden. Het is ook mogelijk om gebruik van lidwoorden (en aanwijzende voornaamwoorden) te verkennen met inbegrip van de sequenties met een adjectief. Daarvoor kan deze zoekreeks worden gebruikt: [pos='lid.*'] []{0,2} [pos='n.*']. De notatie middenin geeft weer dat een lidwoord op maximaal 2 posities afstand gevolgd kan worden door een nomen. (In de zoekopdracht [pos='lid.*'] [pos='adj.*']{0,2} [pos='n.*'] wordt nader gespecificeerd dat op die tussenpositie een adjectief kan staan).

Met de verbuiging van adjectieven ligt het gecompliceerder. De zoekopdracht: "een" [pos="adj.*"]+ [pos="n.*zijd.*"] geeft toegang tot alle onbepaalde contexten waarin een zijdig naamwoord voorafgegaan wordt door een adjectief. De meeste fouten komen we in ons corpus tegen wanneer we in precies deze context onzijdige naamwoorden selecteren. Dan vinden we naast correcte vormen zinsdelen als 'een rode boek'. Wanneer we de bepaalde lidwoorden gebruiken zien we dat in de taalgestoorde tweetalige groep vrijwel exclusief het lidwoord 'de' wordt gebruikt. We kunnen dat beeld ook controleren door de sequenties detN te onderzoeken. In dezelfde groep vinden we dan meer gebruik van 'het', maar bijna exclusief bij verkleinwoorden.

Met de zoektermen [pos='adj.*met.*'] en [pos='adj.*zonder.*'] kunnen we voorkomende gevallen zoeken van respectievelijk adjectieven met en zonder buigings-e.

- (3) *We zouden per kind per stadium de gevallen van gebruik van diverse persoonsvormen uit de corpora willen kunnen filteren en kwantificeren, liefst per type zin: hoofdzin, bijzin, en van de hoofdzinnen nog hoofdzinnen met inversie en zonder inversie.*
- Vervoeging naar enkelvoud en meervoud kan worden gezocht met deze sequenties: [pos='ww.*pv.*tgw.*ev.*'] en [pos='ww.*pv.*tgw.*mv.*']. Probleem bij de tweede zoeksequentie is wel dat ook zinnen worden gevonden van het type 'ik ga water drinken'. Daar wordt immers een vorm op *-en* aangetroffen na een meervoudig onderwerp. De tagging maakt geen onderscheid tussen 1^e, 2^e en 3^e persoon. Het is echter wel mogelijk gebruik te maken van het feit dat pronomina wel als zodanig zijn gemarkeerd, zoals: VNW(pers,pron,nomin,vol,1,ev). Het samen voorkomen van voornaamwoord en (voor tijd en getal) gemarkeerde werkwoorden kan zo de zoekset verkleinen.
- Bij het FESLI-project is gebruik gemaakt van het taggen van constituenten. Het syntactische onderscheid tussen hoofd- en bijzin zit op syntactisch niveau en valt dus buiten de mogelijkheden van een analyse op *parts of speech*.

- (4) *Hoe vinden we uitingen waarin in plaats van een vervoegd lexicaal werkwoord een vorm van 'gaan' wordt gebruikt?*
- Voor deze vraag is het nodig om een nog niet besproken optie te kiezen, dat is: op lemma zoeken ([lemma="gaan"]). Zo worden alle vormen van in dit geval (en het meest prototypisch) 'gaan' gevonden. De vormen van het lexicale werkwoord worden zo eveneens gevonden [bij voorbeeld: lemma="lezen"].