

Linguïsten besparen tijd met CLARIAH-tool

Online tool SASTA analyseert taal

Automatische grammaticale analyse is niet alleen nuttig voor de wetenschap. Ook de klinische praktijk kan deze tool goed gebruiken, zo laat het CLARIAH-project SASTA zien.

Erica Renckens

“Voor een goed behandelplan is een grammaticale analyse van de afwijkende taal essentieel”, vertelt Margo Zwitserlood, die als klinisch linguïst en onderzoeker in het Pento Audiologisch Centrum Amersfoort vooral jonge kinderen met taal- en spraakproblemen ziet. “We nemen dan een gesprek met het kind op, dat we transcriberen en vervolgens analyseren. Daar ben je in totaal vier tot vijf uur mee bezig. Logopedisten hebben daar onvoldoende tijd voor of het ontbreekt hen aan kennis.”

Maatschappelijke impact

In het project SASTA werkt zij daarom met collega's van de Vereniging voor Klinische Linguïstiek en CLARIAH-onderzoekers aan

een programma voor de automatische analyse van spontane taal. Dit programma is niet alleen bedoeld voor kindertaal, maar ook voor afwijkende taal bij volwassenen met afasie. “De basis van SASTA bestaat uit bestaande onderdelen uit de CLARIAH-infrastructuur, zoals de Alpino-parser en de GrETEL-zoekapplicatie voor treebanks”, vertelt Jan Odijk, hoogleraar taal- en spraaktechnologie aan de Universiteit Utrecht. De parser ontleedt de zinnen, waarna met GrETEL in de grammaticale structuren gezocht kan worden. “Het is mooi om te zien hoe deze technieken ook maatschappelijke impact kunnen genereren.”

Eén druk op de knop

SASTA maakt met één druk op de knop een complexiteitsanalyse van de transcriptie. Daarin staat onder andere hoe vaak verschillende woordsoorten en zinsconstructies voorkomen, en hoe dit zich verhoudt tot normaal taalgebruik. De uitkomst van de analyse wordt verwerkt in de scoreformulieren van drie veelgebruikte diagnostische methodes. “Dat scheelt al heel veel werk, geweldig”, aldus Zwitserlood.

In een vergelijking tussen automatisch en



Logopediste Henriette Tissingh en de zevenjarige Noud Thus tijdens een afspraak. Credits: logopediepraktijk Puur Logopedie, Harderwijk

manueel uitgevoerde analyses scoort SASTA erg hoog, tussen de 75 en 90%. Odijk: “Het systeem vindt soms woordsoorten en zinsconstructies die de linguïsten zelf niet uit de data haalden; het werkt dan dus beter dan de mens.” Wel heeft SASTA nog moeite met afwijkende taal: het probeert er correct Nederlands in te zien. Zo analyseert het programma een zin met een congruentiefout als twee zinnen en ziet hij ‘stukkies’ als een samenstelling van ‘stuk’ en ‘kies’. Odijk: “De foutenanalyse doen we daarom nu nog handmatig,

maar de betrokken onderzoekers werken aan de automatisering hiervan.” De transcriptie zal voorlopig ook nog handmatig moeten gebeuren, omdat kinderspraak moeilijk automatisch te herkennen is, zeker als deze afwijkend is. Zwitserlood: “Maar ook daar zijn wetenschappers volop mee bezig, onder andere in Nijmegen. Alle tijdswinst is welkom!”

Een voorlopige versie van SASTA is momenteel online beschikbaar voor linguïsten via: sasta.hum.uu.nl

Data Nationaal Vrijheidsonderzoek beschikbaar

LISS panel meet draagvlak Dodenherdenking en Bevrijdingsdag

Uit het Nationaal Vrijheidsonderzoek blijkt dat een ruime meerderheid van de Nederlandse bevolking de Dodenherdenking op 4 mei en Bevrijdingsdag op 5 mei belangrijk vindt.

Marije Oudejans

Het Nationaal Vrijheidsonderzoek van Nationaal Comité 4 en 5 mei bestaat al sinds 2001. Voor de jaarlijkse meting wordt vanaf 2020 gebruik gemaakt van het LISS panel van CentERdata. Aan dit onderzoek, waarin wordt gevraagd naar de ervaringen van een jaar eerder, hebben ruim 1.700 personen meegedaan. Dezelfde personen worden de komende jaren gevolgd.

Beleving van vrijheid

In het onderzoek komt het thema ‘draagvlak voor 4 en 5 mei’ ieder jaar aan bod. Daarnaast worden om het jaar andere thema's belicht. Zo werd in de meting van maart 2020 ingezoomd op ‘beleving van vrijheid’. Waar denken men-



sen aan bij het begrip vrijheid? Welke gebeurtenissen uit het verleden zijn bepalend voor de ideeën die iemand heeft over vrijheid? De Tweede Wereldoorlog blijkt de gebeurtenis die met afstand het meest bepalend is geweest voor de ideeën die men heeft over

vrijheid en onvrijheid.

Samenhang in data

88% van de Nederlandse bevolking vindt de Nationale Herdenking op 4 mei belangrijk of heel belangrijk. De groepen die hierin wat af-

De afbeelding toont welke vijf onderdelen van 5 mei de respondenten (heel erg) aanspreken. Alle resultaten staan in het onderzoeksrapport van het Nationaal Comité 4 en 5 mei, beschikbaar via 4en5mei.nl. Credits: 4en5mei.nl/onderzoek/nationaal_vrijheidsonderzoek/nationaal-vrijheidsonderzoek-2020

wijken, maar waarvan ook de meerderheid de herdenking (heel) belangrijk vindt, zijn jongeren (75%) en mensen met een niet-westerse migratieachtergrond (70%). Maar liefst 89% hield op 4 mei 2019 om 20:00 uur twee minuten stilte tijdens de Dodenherdenking. Verder blijkt uit het onderzoek dat 4 en 5 mei voor veel mensen met elkaar samenhangen, waarbij de Dodenherdenking voor de meesten net iets belangrijker is dan Bevrijdingsdag.

De door het LISS panel verzamelde data zijn beschikbaar via: dataarchive.lissdata.nl

LISS panel

Het LISS panel is met 5000 huishoudens en zo'n 7500 panelleden een representatieve afspiegeling van de Nederlandse bevolking en sinds 2007 hét panel voor wetenschappelijk, maatschappelijk en beleidsrelevant onderzoek. Joris Mulder, sinds 2020 LISS coördinator: “Representativiteit, datakwaliteit, innovatie en een optimale respons staan bij ons centraal. Hiervoor werken we samen met onder andere ODISSEI, CBS, DANS en SURF. Met regelmaat worden innovatieve experimenten in het LISS panel afgenomen en steeds vaker wordt survey onderzoek gecombineerd met data science. Ook nemen we deel aan het ODISSEI Portal en worden nieuwe projecten voor studies met sensors en wearables gepland.” (EdC) lissdata.nl

SWORD voor LISS Data Archive en DANS

Het LISS Data Archive bevat gegevens over het leven van Nederlanders en hun mening over verschillende onderwerpen, verzameld in een groot aantal onderzoeken onder de leden van het LISS panel. Voor duurzame opslag voor langere termijn worden de data ook gearchiveerd bij DANS.

Sinds kort is er een geautomatiseerde interface tussen het LISS Data Archive en DANS. Met deze zogenaamde SWORD-interface (Simple Web-service Offering Repository Deposit) worden de data grotendeels automatisch doorgestuurd naar DANS. Zo is dubbel invoerwerk niet meer nodig en worden de data sneller doorgegeven. (MdB) dans.knaw.nl