



Signaal Digitaal

Significant voor de
professionele hulpverlener

2025 - nummer 4

Verschenen op 11 december 2025

Signaal Digitaal is het digitale tijdschrift van Sig vzw.
Meer info op www.sig-net.be > Signaal Digitaal

Redactie

Geert Andries
Dieter Baeyens
Henri De Vroey
Greetje Desnerck
Hilde Roeyers
Herman Van Hove
Petra Warreyn
Inge Zink



INCLUSIE IN-ZICHT

De werking van Sig wordt ondersteund door het Wetenschappelijk Netwerk, bestaande uit meer dan 40 leden verbonden aan universiteiten en hogescholen. De volledige namenlijst vindt u op www.sig-net.be > Wetenschap

Redactiesecretariaat

Tijdschrift Signaal Digitaal > signaaldigitaal@sig-net.be
p/a Sig vzw, Pachthofstraat 1, 9308 Gijzegem (Aalst) (B)

APROCSA-NL: een tijdsefficiënt, betrouwbaar en valide instrument om natuurlijke spraak bij personen met afasie in kaart te brengen

Mara Barberis ⁽¹⁾⁽²⁾⁽³⁾, Marianne Casilio⁽⁴⁾,
Stephen M. Wilson⁽⁵⁾ en Maaïke Vandermosten ⁽¹⁾⁽²⁾⁽³⁾

(1) Experimentele Otorhinolaryngologie (ExpORL), Departement Neurowetenschappen, KU Leuven

(2) Leuven Brain Institute, KU Leuven

(3) Leuven Interdisciplinary Language Institute, KU Leuven

(4) Department of Hearing and Speech Sciences, Vanderbilt University Medical Center, Nashville, TN, USA

(5) School of Health and Rehabilitation Sciences, University of Queensland, Brisbane, QLD, Australië

Personen met afasie hebben vaak moeite met het produceren van natuurlijke spraak, wat de dagdagelijkse communicatie ernstig kan belemmeren. Daarom is het assessment van natuurlijke spraak bij personen met afasie van cruciaal belang. Er zijn momenteel echter geen tijdsefficiënte, comprensieve instrumenten voor natuurlijke spraak beschikbaar voor het Nederlands.

Dit artikel bespreekt de vertaling en validatie van APROCSA-NL, de Nederlandstalige versie van het instrument Auditory-Perceptual Rating of Connected Speech in Aphasia. APROCSA-NL werd gevalideerd op basis van een proefgroep van 58 personen met chronische afasie ten gevolge van een beroerte. Hun natuurlijke spraakstalen werden beoordeeld door drie ervaren en drie onervaren beoordelaars, zodat de invloed van ervaring op de interbeoordelaarsbetrouwbaarheid van de APROCSA-NL nagegaan kon worden. Daarnaast werd ook onderzocht hoe de ernst van de afasie en de duur van het spraakstaal de interbeoordelaarsbetrouwbaarheid beïnvloeden. Tot slot werd ook de validiteit van de APROCSA-NL nagegaan en werden de onderliggende constructen van natuurlijke spraak onderzocht via factoranalyse.

De meeste APROCSA-NL features hadden een goede tot uitstekende interbeoordelaarsbetrouwbaarheid, en bleken valide metingen van natuurlijke spraak. Bijkomend toonden we aan dat APROCSA-NL ook gebruikt kan worden bij kortere natuurlijke spraaksamples van minder dan drie minuten en in gevallen van ernstige afasie. We vonden vier onderliggende constructen van natuurlijke spraak (Onduidelijkheid, Agrammatisme, Spraak-motorisch en Anomie) die 65% van de variantie in natuurlijke spraak verklaarden. APROCSA-NL bleek daarnaast ook een tijdsefficiënte methode, met een gemiddelde scoringsduur tussen de negen en elf minuten.

We besluiten dat APROCSA-NL een tijdsefficiënt, comprensief, betrouwbaar en valide meetinstrument is om de natuurlijke spraakvaardigheden bij personen met afasie in kaart te brengen.

1. Inleiding

Personen met afasie ondervinden vaak moeilijkheden bij het produceren van natuurlijke spraak, wat de alledaagse communicatie ernstig kan bemmeren. Dit kan leiden tot frustratie, een verminderde levenskwaliteit, psychosociale stress en sociale isolatie (Stark, 2023). Toch zijn afasie assessments vaak sterk gericht op geïsoleerde taaltaken met losse fonemen, woorden of zinnen (Leaman & Edmonds, 2021), die een beperkte ecologische validiteit hebben. Dit kan leiden tot discrepanties tussen de klinische beoordeling en de eigenlijke communicatieve vaardigheden die de persoon met afasie in het dagelijks leven ervaart (Cavanaugh & Haley, 2020; Fromm e.a., 2017; Holland e.a., 2019; Kim e.a., 2022; Lesser & Algar, 1995; Stark & Fukuyama, 2021; Wallace & Kimelman, 2013). Daarom pleiten klinici en onderzoekers ervoor om systematisch natuurlijke spraaktaken te includeren bij het beoordelen van afasie (Azios e.a., 2021; Dietz & Boyle, 2018).

Er werden al verschillende natuurlijke spraakfeatures gedefinieerd in de afasiologie (Bryant e.a., 2016), maar de psychometrische eigenschappen van deze features zijn slechts beperkt onderzocht (Boyle, 2014, 2015; Brookshire & Nicholas, 1994; Pritchard e.a., 2018; Stark & Dalton, 2024). Bijkomend wordt natuurlijke spraakanalyse nog te beperkt toegepast in de praktijk door de hoge workload (Bryant e.a., 2019; Hallowell, 2016; Stark e.a., 2021), wat typisch het manueel transcriberen en analyseren van natuurlijke spraak inhoudt. Dit proces kan tot één uur duren per minuut aan natuurlijke spraak (Armstrong e.a., 2007). Als reactie hierop werden *transcription-less* instrumenten ontwikkeld (zie bv. Armstrong e.a., 2007; Casilio e.a., 2019; Kim & Wright, 2020; Ruiten e.a., 2022; Stark & Dalton, 2024). Deze zijn bruikbaar in de klinische praktijk aangezien de evaluatie van natuurlijke spraak op basis van het beluisteren van de audio-opname gebeurt, zonder dat er een transcriptie gemaakt moet worden.

Om tegemoet te komen aan de nood voor een klinisch inzetbaar, comprehensief instrument om natuurlijke spraak bij personen met afasie te beoordelen, ontwikkelden Casilio en collega's

in 2019 de *Auditory-Perceptual Rating of Connected Speech in Aphasia (APROCSA)*. Het instrument bestaat uit 27 natuurlijke spraakfeatures die elk beoordeeld worden op een vijfpuntschaal, variërend van 0 (niet aanwezig of binnen het bereik van gezonde, oudere sprekers) tot 4 (ernstig, kenmerk is bijna altijd aanwezig). Wat de scoring van APROCSA-NL onderscheidt van andere afasie-instrumenten, is dat er aan een bepaalde score niet alleen een ernstgraad (bv. mild, matig, ernstig), maar ook een beschrijving van deze ernstgraad gekoppeld wordt, wat de logopedist meer houvast geeft bij het scoren van de natuurlijke spraakfeatures. De 27 natuurlijke spraakfeatures zijn onderverdeeld in zeven categorieën: (1) ophalen van woorden, (2) selectie van woorden en klanken, (3) grammaticale constructies, (4) tempo en timing, (5) zelfcorrecties, (6) duidelijkheid en (7) diagnostisch. Tabel 1 geeft een overzicht van de features per categorie.

Voor het Nederlands zijn er slechts een beperkt aantal natuurlijke spraakinstrumenten voor afasie beschikbaar. Bovendien hebben ze vaak beperkingen zoals een zeer arbeidsintensieve scoringsprocedure of een beperkte representatie van de diverse aspecten van natuurlijke spraak. Daarnaast ontbreekt het vaak aan een uitgebreide psychometrische validatie.

De subtest Situatieplaatbeschrijving van de Nederlandstalige bewerking van de Comprehensive Afasia Test (CAT-NL) (Visch-Brink e.a., 2015) omvat een auditief-perceptuele beoordeling van de nauwkeurigheid en adequaatheid van de informatie, het tempo en de vloeiendheid, de grammaticale variatie en correctheid en een algemene beoordeling. Er is echter weinig houvast om aan deze aspecten een score toe te kennen. Bovendien zijn er geen psychometrische gegevens beschikbaar voor deze subtest van de CAT-NL. Recent werd de conceptuele scoring van de situatieplaat toegevoegd (Vandenborre e.a., 2021), waarbij een objectieve, kwantitatieve maat wordt bekomen voor het produceren van sleutelconcepten, het leggen van causale verbanden en het totaal aantal woorden. De interbeoordelaarsbetrouwbaarheid van deze maten is uitstekend ($ICC_{\text{sleutelconcepten}} = 0.94$, $ICC_{\text{aantal woorden}} = 0.99$, $ICC_{\text{causaal}} = 0.92$) (Vandenborre e.a., 2018).

Tabel 1: Overzicht van de APROCSA-NL features per APROCSA-NL categorie

APROCSA-NL categorie	APROCSA-NL feature
Ophalen van woorden	Anomie Onvolledige uitingen Lege spraak Circumlocutie (omschrijving)
Selectie van woorden en klanken	Semantische parafasieën Fonematische parafasieën Neologismen Jargon Perseveratie Stereotypen en automatismen
Grammaticale constructie	Korte en vereenvoudigde uitingen Omissie van functiewoorden Omissie van het gebonden morfeem Paragrammatisme
Tempo en timing	Pauze tussen uitingen Pauze binnen uitingen Aarzelend en moeizaam Verminderd spreektempo
Zelfcorrectie	Hernemingen Valse start Conduite d'approche
Duidelijkheid	Doelwoord of -uiting onduidelijk Betekenis onduidelijk Off topic
Diagnostisch	Expressieve afasie Spraakapraxie Dysartrie Algemene communicatiebeperking

De Nederlandstalige bewerking van de Main Concept Analysis (MCA) (Criel e.a., 2021) kwantificeert de effectiviteit en efficiëntie van de informatieoverdracht op basis van de aanwezigheid, accuraatheid en volledigheid van de hoofdconcepten ('main concepts') die geproduceerd worden tijdens vier sequentiële situatieplaatbeschrijvingen. De MCA-scores toonden een goede tot uitstekende interbeoordelaarsbetrouwbaarheid (Criel e.a., 2025).

Bij de ANTAT vervolgens wordt natuurlijke spraak gemeten via de respons van de persoon

met afasie op een alledaagse situatie (bv. een afspraak verzetten). De verbale responsen worden beoordeeld op een vijfpuntschaal met betrekking tot de verstaanbaarheid en de begrijpelijkheid van de boodschap. Van deze rating werd al aangetoond dat ze betrouwbaar en valide is (Blomert e.a., 1994).

Onlangs werd de Content Unit (CU) scoring toegevoegd aan de ANTAT. Van deze CU-scoring werd eveneens aangetoond dat deze betrouwbaar en valide is (Ruiter e.a., 2022).

Een belangrijke kanttekening bij de zonet besproken tests is dat ze slechts beperkte aspecten van natuurlijke spraak meten, met een sterke focus op informatieoverdracht (cf. sleutelconcepten in de situatieplaat van de CAT- NL, Main Concepts in de MCA, Begrijpelijkheid / Content Units in de ANTAT).

De Analyse voor Spontane Taal bij Afasiepatiënten (ASTA) (Boxum e.a., 2019; Van Der Scheer e.a., 2011) daarentegen is een meer comprehensieve methode die 19 natuurlijke spraakmetingen overheen verschillende taaldomeinen in kaart brengt. De interbeoordelaarsbetrouwbaarheid voor negen ASTA features varieerde van redelijk tot uitstekend (Van Der Scheer e.a., 2011). De ASTA is echter zeer tijdsintensief door het transcriberen en uitgebreid linguïstisch analyseren van het taalstaal, en wordt bijgevolg weinig ingezet in de klinische praktijk.

Het doel van deze studie is om de Engelstalige APROCSA te vertalen naar het Nederlands (APROCSA-NL) en de interbeoordelaarsbetrouwbaarheid en concurrente validiteit ervan na te gaan, zowel voor ervaren als niet ervaren beoordelaars.

Op deze manier draagt de studie bij aan het beschikbaar maken van een comprehensief en tijdsefficiënt instrument om natuurlijke spraak bij personen met afasie in kaart te brengen. Verder dragen we ook bij aan de bredere APROCSA-onderzoekslijn, aangezien we de invloed van afasie-ernst en sample duur op de betrouwbaarheid van de APROCSA-scores nagaan. Tot slot onderzoeken we de onderliggende constructen van natuurlijke spraak aan de hand van de APROCSA-scores in een sample van 58 personen met afasie.

2. Methode

2.1 Vertaling

APROCSA werd vertaald van het Engels naar het Nederlands volgens de principes van *forward and backward translation*, wat de meest aanbevolen aanpak is om gezondheidsgerelateerde instrumenten te vertalen (Beaton e.a.,

2000; Sousa & Rojjanasrirat, 2011; Wild e.a., 2005). Deze aanpak vereiste twee onafhankelijke vertalers: de eerste (auteur M.B.) vertaalde APROCSA naar het Nederlands, de tweede (een tweetalige Engels-Nederlands spreker) vertaalde deze versie terug naar het Engels. Deze laatste versie werd gecheckt met de originele versie, ontwikkeld door Casilio en collega's.

Discrepancies tussen deze versies waren voornamelijk te wijten aan logopedisch jargon, waarmee de tweetalige Engels-Nederlands spreker niet vertrouwd was. Auteurs M.B. en M.C. bespraken deze gevallen en er werden nog kleine aanpassingen doorgevoerd aan de Nederlandse vertaling van APROCSA.

2.2 Deelnemers

Aan deze studie participeerden 58 personen met chronische afasie ten gevolge van een beroerte (≥ 6 maanden post onset). Personen werden gerekruteerd via de stroke unit van het UZ Leuven en via zelfstandige logopedisten.

De deelnemers werden gediagnosticeerd met afasie in de acute fase door een logopedist, en scoorden op het moment van deelname aan de studie afwijkend op de ScreeLing (Visch-Brink e.a., 2010) of Nederlandse Benoem Test (NBT) (Van Ewijk e.a., 2018) of rapporteerden communicatieve problemen volgens de Communicatieve Effectiviteits Index (CETI) (Lomas e.a., 1989).

Personen met afasie moesten geen specifiek type afasie hebben om geïnccludeerd te worden, aangezien we een zo representatief mogelijk sample van de heterogene populatie van personen met afasie wilden verzamelen.

Deelnemers werden geëxcludeerd in het geval van een comorbide psychiatrische of neurodegeneratieve aandoening. Tabel 2 geeft een demografische en klinische karakterisering van de deelnemersgroep weer. Alle deelnemers gaven schriftelijke toestemming om deel te nemen aan de studie. Deze studie werd goedgekeurd door het Medisch-Ethisch Comité Onderzoek van het UZ Leuven en de KU Leuven (B32220-1731747).

Tabel 2: Demografische en klinische karakteristieken van de deelnemersgroep

Leeftijd	67.1 ± 13.4 (32-95) jaar
Geslacht	27 M; 31 V
Handvoorkeur	41 R; 14 L; 3 Am
Opleidingsniveau	7 LO; 27 MO; 16 H; 8 U
Tijd post onset	38.7 ± 58.2 (6-360) maanden
Afasie ernst ¹	58.9 ± 12.1 (19-72)
Woordvinding ²	208.4 ± 76.6 (7-274)
Communicatieve effectiviteit ³	6.6 ± 1.8 (2.7-10.0)
Dysartrie ⁴	41.38%
Spraakapraxie ⁴	36.21%

M = Man, V = Vrouw, R = Rechtshandig, L = Linkshandig, Am = Ambidexter, LO = Lager onderwijs, MO = Middelbaar onderwijs, H = Hogeschool, U = Universiteit. ¹ Gebaseerd op de ScreeLing (Visch-Brink e.a., 2010). Totalscores lager dan 68 wijzen op afasie, maximumscore = 72. ² Gebaseerd op de Nederlandse Benoem Test (Van Ewijk e.a., 2018). Totalscores lager dan 255 wijzen op woordvindingsproblemen, maximumscore = 276. ³ Gebaseerd op de Communicatieve EffectiviteitsIndex (CETI) (Lomas e.a., 1989). Scores op de zestien items werden gemiddeld tot een score op tien, waarbij lagere scores een lagere communicatieve effectiviteit weerspiegelen. ⁴ Binaire classificatie (aanwezig/afwezig) door een logopedist.

2.3 Natuurlijke spraaktaak

Natuurlijke spraak werd uitgelokt door het stellen van twee open vragen, nl. ‘Wat heb je afgelopen weekend gedaan?’ en ‘Kan je me iets vertellen over de dag dat je je beroerte gekregen hebt?’. Indien de deelnemers niet antwoordden binnen de eerste tien seconden, herhaalde de testleider de vraag, of werd een gerelateerde vraag gesteld, zoals ‘Wat is je eerste herinnering na de beroerte?’. Dit is in overeenstemming met het protocol voor het uitlokken van spontane spraak van AphasiaBank (MacWhinney e.a., 2011). De duur van de natuurlijke spraak-samples varieerde van 31 tot 937 seconden ($M = 236.3$ s, $SD = 185.2$).

In tegenstelling tot de originele APROCOSA-studie, stelden we geen minimum criterium van vijf minuten natuurlijke spraak, aangezien één van onze doelen was om na te gaan of APROCOSA ook bruikbaar is voor personen met ernstige afasie, waarbij het vaak niet haalbaar is om vijf minuten natuurlijke spraak te verzamelen.

Een tweede verschil met het proefopzet van de originele APROCOSA-studie heeft betrekking op de opnamemodaliteit. In de originele studie

maakte men zowel audio- als video-opnames, terwijl in de huidige studie enkel audio-opnames beschikbaar waren. Deze audio-opnames werden in een rustige ruimte bij de deelnemer thuis gemaakt met een BOYA BY-M1 Universal Lavalier Microfoon en een Philips DVT 1120 VoiceTracer met een sample rate van 16 kHz en een bitdiepte van 16.

2.4 Beoordelaars

Twee afzonderlijke groepen van Nederlandstalige beoordelaars scoorden de APROCOSA-NL voor het sample van 58 personen met afasie. De eerste groep bestond uit drie onervaren beoordelaars, meer specifiek, eerstejaars masterstudenten logopedie die theoretische achtergrond hadden over neurogene communicatiestoornissen, maar nog geen klinische ervaring opgedaan hadden. De tweede groep bestond uit drie ervaren raters. De eerste rater was logopedist van opleiding en werkzaam als onderzoeker, met vier jaar ervaring in het transcriberen en analyseren van natuurlijke spraak bij personen met afasie. De tweede en derde ervaren raters waren logopedisten met zeven tot negen jaar klinische ervaring en bijzondere expertise in neurogene communicatiestoornissen, werkzaam in zowel acute als chronische settings.

2.5 Beoordelingsprocedure

Voor het beoordelen van de eigenlijke validatiesamples scoorden de ervaren beoordelaars elk vijf kalibratiesamples. Deze kalibratiesamples bestonden uit audio-opnames van de mondelinge beschrijving van de situatieplaat van de CAT-NL. De samples weerspiegelden diverse afasieprofielen en waren specifiek geselecteerd, zodat de verschillende APROCSA-NL features aan bod zouden komen. Nadat de ervaren beoordelaars de kalibratiesamples afzonderlijk gescoord hadden, werden deze beoordelingen besproken en werd er consensus bereikt over de vijf kalibratiesamples.

De onervaren beoordelaars kregen elk een anderhalf uur durende training met uitleg over het doel van APROCSA-NL, uitgebreide toelichting over de APROCSA-NL features en richtlijnen voor het scoren. Na deze trainingssessie oefenden de studenten onafhankelijk met de vijf kalibratiesamples. Een week later was er een tweede trainingssessie om hun ratings te vergelijken met de consensusratings. Deze procedure zorgde ervoor dat alle beoordelaars gekalibreerd waren volgens dezelfde consensusscores.

De beoordelaars scoorden elk de 58 samples van personen met afasie over een tijdspanne van twee maanden. Er werd aan de beoordelaars gevraagd om niet langer dan één uur aaneensluitend te scoren om vermoeidheidseffecten te elimineren. Het toekennen van de APROCSA-NL scores gebeurde op dezelfde manier als in de originele APROCSA-studie: beoordelaars luisterden één keer naar het sample en scoorden vervolgens het merendeel van de features. Nadien was er de mogelijkheid om nogmaals te luisteren indien er over bepaalde features nog onduidelijkheden waren. Beoordelaars werden geadviseerd om notities te maken tijdens het luisteren van de sample. Daarnaast werd – om het tijdsefficiënte karakter van APROCSA-NL aan te tonen – aan iedere beoordelaar gevraagd om bij te houden hoe lang ze over iedere scoring deden.

2.6 Interbeoordelaarsbetrouwbaarheid

Intraclass correlatiecoëfficiënten (ICC's) werden berekend als maat voor interbeoordelaarsbetrouwbaarheid (IBB). ICC (A,k), die de absolute overeenstemming schat van metingen die gemiddelden zijn van k onafhankelijke metingen, werden berekend voor tweerichtingsmodellen met behulp van het *irr package* in R (versie 4.4.1).

We onderzochten kwalitatief of het ervaringsniveau een invloed had op de ICC's van de APROCSA-NL features. ICC's werden als volgt geïnterpreteerd: slecht ($r < 0,40$), redelijk ($0,40 \leq r < 0,60$), goed ($0,60 \leq r < 0,75$) en uitstekend ($r \geq 0,75$) (Cicchetti, 1994). Vervolgens werd de invloed van de duur van de natuurlijke spraaksamples en de ernst van de afasie op de betrouwbaarheid van APROCSA-NL-scores onderzocht aan de hand van de ICC's van de ervaren groep, ervan uitgaande dat de ervaren groep accuratere APROCSA-NL-scores zou opleveren in vergelijking met de groep onervaren studenten.

Casilio en collega's analyseerden in de originele studie natuurlijke spraaksamples van minimaal vijf minuten, aangezien dit de aanbevolen sampleduur is door Boles en Bombard (1998) om communicatieve effectiviteit bij afasie te kunnen beoordelen. Brookshire en Nicholas (1994) suggereren daarentegen dat samples van 300 à 400 woorden voldoende zouden moeten zijn om betrouwbare informatie over de taalvaardigheden bij afasie te verkrijgen. Aangezien personen met afasie in onze studie een gemiddeld spreektempo van 110 woorden per minuut hadden, zou hun sample ongeveer drie minuten lang zijn om een sample tussen 300 en 400 woorden te bekommen. Daarom evalueerden we de invloed van sampleduur op de IBB van de APROCSA-NL features, gebruikmakend van zowel de drie als vijf minuten cutoff. We verdeelden de participanten bijgevolg in drie groepen: (1) sampleduur van minder dan drie minuten ($n = 29$), (2) sampleduur tussen drie en vijf minuten ($n = 14$) en (3) sampleduur van meer dan vijf minuten ($n = 15$).

Tabel 3: Gemiddelde, standaarddeviatie en bereik van de scores op de ScreeLing en de Nederlandse Benoem Test (NBT) voor de ernstige, matige en milde afasiegroepen

	Ernstig (n = 5)	Matig (n = 9)	Mild (n = 35)
ScreeLing	30.2 ± 9.67 (19-41)	47.2 ± 3.67 (42-53.5)	63.1 ± 4.9 (54-71)
NBT	64.6 ± 79.9 (7-179)	128.6 ± 93.6 (17-233)	234.7 ± 26.7 (141-269)

Om de invloed van de afasie-ernst op de IBB na te gaan, werden drie subgroepen gedefinieerd op basis van de scores op de ScreeLing en de Nederlandse Benoem Test (Tabel 3).

De eerste groep (n = 5) bestond uit personen met ernstige afasie met ScreeLing-scores < 41.5. De tweede groep (n = 9) bestond uit personen met matige afasie met ScreeLing-scores tussen 42 en 53.5. De derde groep (n = 35) bestond uit personen met milde afasie met ScreeLing-scores tussen 54 en 67.5 of scores op de Nederlandse Benoem Test lager dan 255 in het geval dat de ScreeLing-score 68 of hoger (en dus niet afwijkend) was. Dit laatste criterium werd toegevoegd om te vermijden dat personen met een zuiver anomische afasie over het hoofd gezien zouden worden, aangezien de ScreeLing veelal receptieve taaltaken bevat.

2.7 Concurrente validiteit

Om de concurrente validiteit van APROCSA-

NL te onderzoeken, werden Benjamini-Hochberg gecorrigeerde Spearman correlaties berekend tussen enerzijds de APROCSA-NL features (d.i., het gemiddelde van de scores van de drie ervaren beoordelaars) en anderzijds zestien kwantitatieve natuurlijke spraakmaten die berekend werden vanuit transcripties (Tabel 4) en vijf scores op gestandaardiseerde afasie assessments (ScreeLing subtests semantiek, fonologie en syntaxis, Nederlandse Benoem Test-score en CETI-score). De APROCSA-NL features werden op voorhand gelinkt aan de kwantitatieve natuurlijke spraakmaten. Aangezien deze geen grammaticale maten bevatten, werden de grammaticale constructiefeatures van APROCSA-NL (korte en vereenvoudigde uitingen, omissie van het gebonden morfeem, omissie van functiewoorden en paragrammatisme) gelinkt aan de subtest syntaxis van de ScreeLing. Op deze manier konden 23 van de 28 APROCSA-NL features gelinkt worden aan een kwantitatieve tegenhanger om de concurrente validiteit na te gaan.

Tabel 4: Transcriptiegebaseerde natuurlijke spraakfeatures, hun operationalisatie en – indien van toepassing – hun APROCSA-NL tegenhanger(s)

Natuurlijke spraakfeatures	Operationalisatie	APROCSI-NL tegenhangers
Aantal woorden	Totaal aantal geproduceerde woorden, inclusief non-woorden, fonematische fouten, herhalingen, minimale responsen, commentaren en stereotypen, in overeenstemming met Boxum e.a. (2013) en Vandenborre e.a. (2018).	/
Neologisme	Non-woorden, meer bepaald woordvormen die geen echte woorden zijn (Casilio e.a., 2019), berekend als een proportie van het totaal aantal woorden.	Neologisme

Natuurlijke spraakfeatures	Operationalisatie	APROCSI-NL tegenhangers
Afgebroken woord	Woorden worden afgebroken na één of twee fonemen; het doelwoord wordt niet geproduceerd. Berekend als een proportie van het totaal aantal woorden.	Onvolledige uitingen, valse start, spraakapraxie
Herneming	Een woord wordt initieel afgebroken, maar het doelwoord wordt uiteindelijk wel geproduceerd (bv. ca-camera). Berekend als een proportie van het totaal aantal woorden.	Valse start, conduite d'approche, spraakapraxie
Onverstaanbaar	Overstaanbare woorden, berekend als een proportie van het totaal aantal woorden.	Doelwoord of -uiting onduidelijk
Fonologische fouten	Foneemdeletie, -insertie, -substitutie of -transpositie (Casilio e.a., 2019; Vermeulen e.a., 1989), berekend als een proportie van het totaal aantal woorden.	Fonematische parafasieën
Semantische fouten	Substitutie van inhoudswoorden voor (on)gerelateerde inhoudswoorden (Casilio e.a., 2019), berekend als een proportie van het totaal aantal woorden.	Semantische parafasieën
Partikels	Totaal aantal discourse partikels, berekend als een proportie van het totaal aantal woorden.	/
Ge vulde pauzes	Ge vulde pauzes (bv. 'uh'), berekend als een proportie van het totaal aantal woorden.	Pauzes binnen uitingen
Woordherhaling	Onmiddellijke herhaling van een eerder gebruikt woord, berekend als een proportie van het totaal aantal woorden.	Herneming
Type Token Ratio (TTR)	Aantal verschillende woorden (types) gedeeld door het aantal woorden (tokens). Hogere waarden betekenen dat de spreker een meer gevarieerde woordenschat gebruikt (Sanborn e.a., 2020).	'Ophalen van woorden' features
Brunet Index (BI)	Meting van lexicale diversiteit die onafhankelijk is van de tekstlengte. Dit wordt berekend als $w(u^{-0.165})$ waarbij w = totaal aantal tokens en u = totaal aantal unieke woordtypes (Parsa e.a., 2021). Lagere waarden duiden op een rijkere woordenschat (Sanborn e.a., 2020).	'Ophalen van woorden' features
Woorden per minuut	Totaal aantal woorden gedeeld door de totale duur in minuten.	'Tempo en timing' features
Fonen per minuut	Totaal aantal fonen gedeeld door de totale duur in minuten.	'Tempo en timing' features
Lange pauzes	Aantal stille segmenten tussen woorden van > 400 ms (Le e.a., 2018; Pakhomov e.a., 2010), gedeeld door het totaal aantal woorden.	Anomie, pauzes tussen uitingen, aarzelend en moeizaam, verminderd spreektempo, spraakapraxie, dysartrie
Korte pauzes	Aantal stille segmenten tussen woorden tussen 150 en 400 ms (Le e.a., 2018; Pakhomov e.a., 2010), gedeeld door het totaal aantal woorden.	Anomie, pauzes binnen uitingen, aarzelend en moeizaam, verminderd spreektempo, spraakapraxie

Transcripties werden verkregen via een Nederlandstalig automatisch spraakherkenningsmodel (Poncelet & Van hamme, 2022), dat getraind werd op 270 uur aan Vlaamse spraak van het Corpus Gesproken Nederlands (Oostdijk, 2000) en gefinetuned op negen uur natuurlijke spraak van achttien cognitief gezonde ouderen. De automatisch gegenereerde transcripties werden manueel gecorrigeerd en gelabeld voor afasiespecifieke fouten zoals semantische fouten, fonologische fouten, neologismen en onvloeiendheden (woordherhaling, afgebroken woord, gevulde pauze).

De kwantitatieve natuurlijke spraakmaten werden zowel manueel als automatisch afgeleid van de transcripties. De manuele maten omvatten het tellen van het aantal afasiespecifieke labels per participant. Voor de automatische maten werd robust Voice Activity Detection (Tan e.a., 2020) gebruikt om de segmenten met stemgeving te extraheren. Nadien werd Montreal Forced Aligner (McAuliffe e.a., 2017) toegepast om timings op woordniveau te bekomen, waaruit features zoals het spreektempo en het aantal pauzes afgeleid werden.

2.8 Onderliggende constructen van APROCSA-NL

Paarsgewijze Benjamini-Hochberg gecorrigeerde Spearman correlaties werden berekend tussen de APROCSA-NL feature scores. Nadien werd een factoranalyse met orthogonale varimax rotatie uitgevoerd op basis van de gemiddelde scores van de ervaren beoordelaars. Factoren werden geëxtraheerd volgens de minimale residuen-methode, gezien de relatief kleine deelnemersgroep en de niet normaal verdeelde data. Het aantal factoren werd bepaald via parallelanalyse. Variabelen met ladingen lager dan 0.4 op een factor of een hoge uniciteit (> 0.75) werden verwijderd uit het factormodel, aangezien dit erop wijst dat deze variabelen geen betekenisvolle bijdrage leveren aan de factorstructuur. De resulterende factorstructuur werd geïnterpreteerd op basis van de ladingen van de overgebleven variabelen.

3. Resultaten

3.1 Tijdsefficiëntie van APROCSA-NL

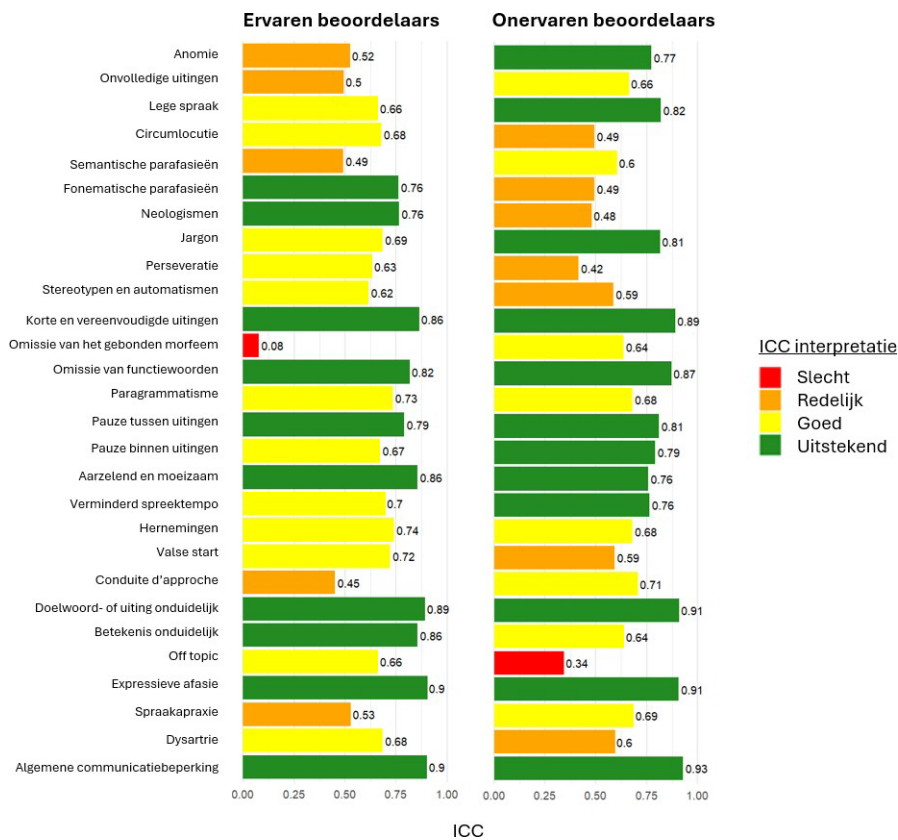
Om de tijdsefficiëntie van APROCSA-NL na te gaan, werd aan de beoordelaars gevraagd hun scoringstijd per deelnemer te registreren. De gemiddelde scoretijd was minder dan vijftien minuten voor zowel de ervaren ($M = 9.1$ minuten, $SD = 3.6$ minuten) als de onervaren beoordelaars ($M = 11.0$ minuten, $SD = 3.7$ minuten), wat in overeenstemming is met de bevindingen uit de originele studie van Casilio en collega's (2019). Zoals verwacht, neemt de scoringstijd significant toe bij toenemende sampleduur (Spearman $\rho = 0.64$, $p < 0.001$).

3.2 Interbeoordelaarsbetrouwbaarheid van APROCSA-NL

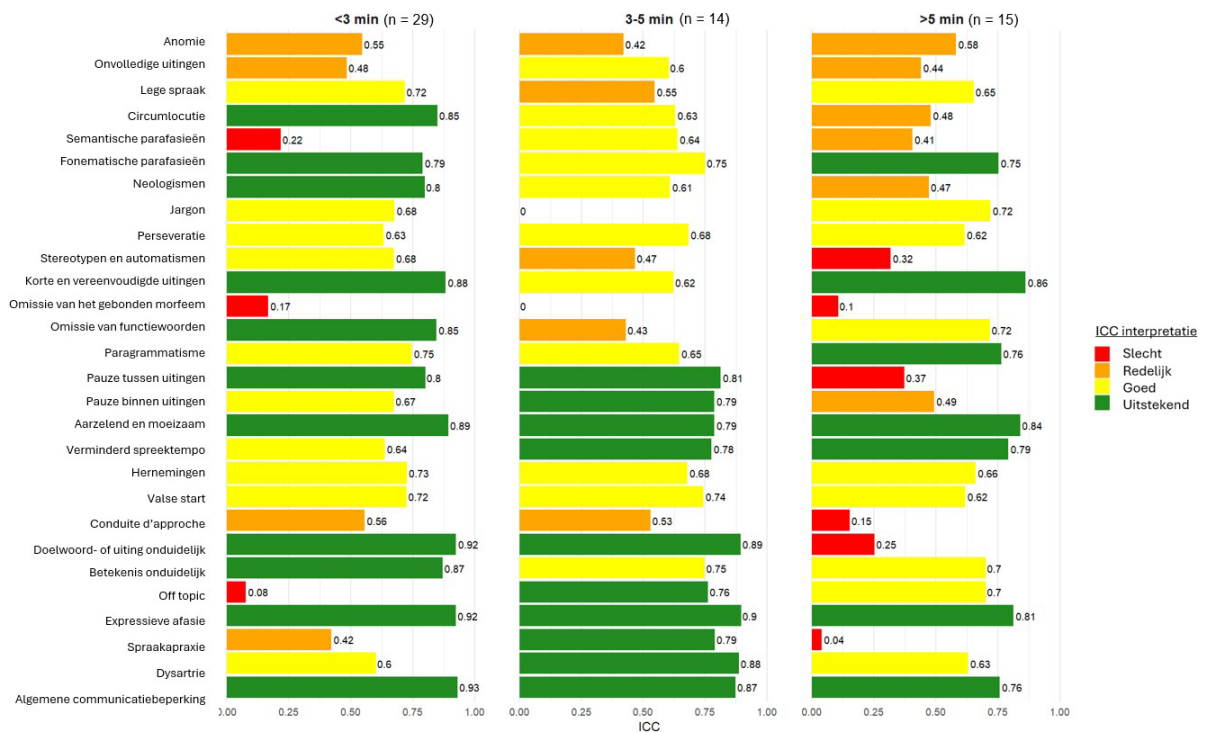
Gebaseerd op ICC (A,k), die de absolute overeenstemming schat van metingen die gemiddelden zijn van k onafhankelijke metingen, behalen de ervaren beoordelaars een uitstekende interbeoordelaarsbetrouwbaarheid (IBB) voor tien features, een goede IBB voor twaalf features, een redelijke IBB voor vijf features en een slechte IBB voor één feature (omissie van het gebonden morfeem) (Figuur 1, kolom 1). De onervaren beoordelaars hadden een uitstekende IBB voor twaalf features, een goede IBB voor acht features, een redelijke IBB voor zeven features en een slechte IBB voor één feature (off topic) (Figuur 1, kolom 2).

Zelfs bij relatief korte natuurlijke spraak-samples van minder dan drie minuten, vertoonden de meerderheid van de features een goede tot uitstekende IBB gebaseerd op ICC (A,k) van de ervaren beoordelaars (Figuur 2). Anderzijds waren er ook enkele features (bv. stereotypen en automatismen, pauzes binnen en tussen uitingen, conduite d'approche en doelwoord of -uiting onduidelijk) die een relatief lagere ICC hadden voor langere (> 5 min) samples in vergelijking met kortere (≤ 5 min) samples.

Figuur 1: Intraclass correlatiecoëfficiënt [ICC (A,k)] en hun interpretatie (kleur-gecodeerd) voor de APROCSA-NL features



Figuur 2: Intraclass correlatiecoëfficiënt [ICC (A,k)] en hun interpretatie (kleur-gecodeerd) voor de APROCSA-NL features, onderverdeeld op basis van sample duur



Figuur 3: Intraclass correlatiecoëfficiënt [ICC (A,k)] en hun interpretatie (kleur-gecodeerd) voor de APROCSA-NL features, onderverdeeld op basis van afasie-ernst



Er konden geen ICC's berekend worden voor circumlocutie, semantische parafasieën en betekenis onduidelijk in de groep 'Ernstig', aangezien er geen variabiliteit in de data was (alle raters waren het eens over score 0 (= afwezig) voor de vijf personen met ernstige afasie).

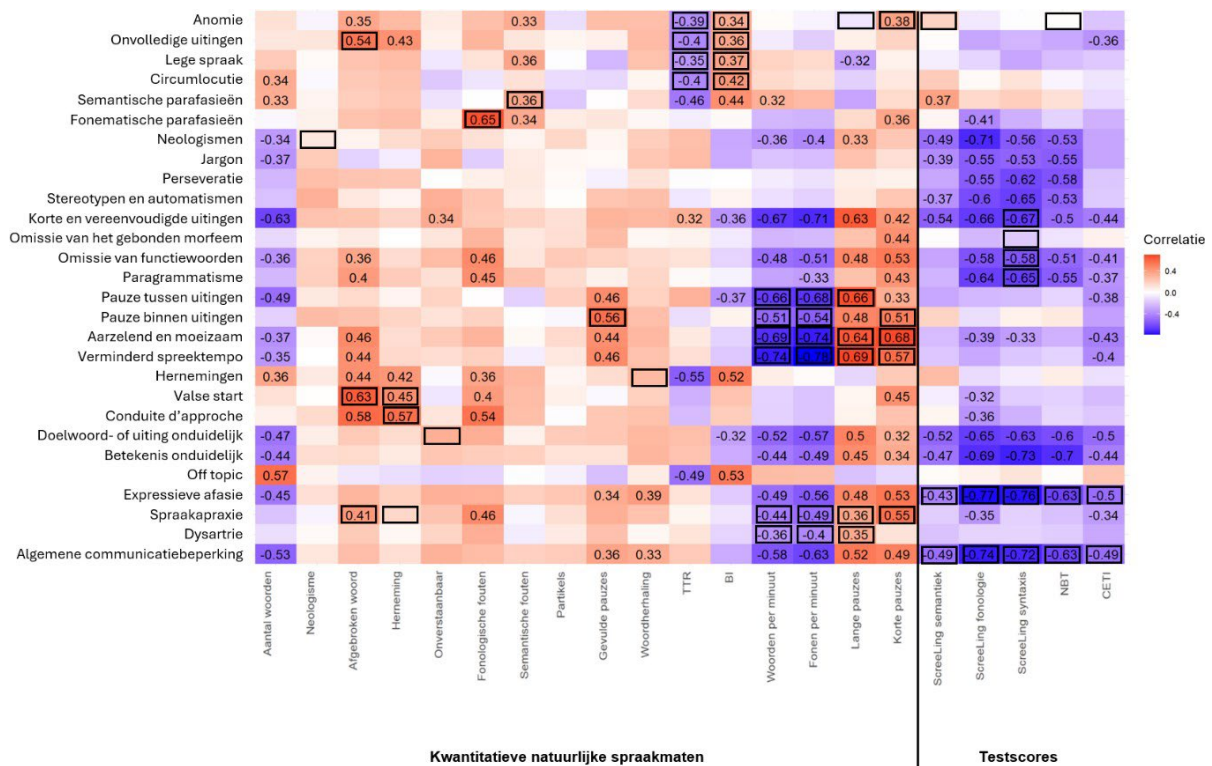
Figuur 3 toont dat niet alle APROCSA-NL features even betrouwbaar zijn voor de verschillende ernstgraden van afasie. Desalniettemin is het een veelbelovende bevinding dat, zelfs voor ernstige afasie, de features jargon, perseveratie, omissie van functiewoorden, pauzes tussen uitingen, aarzelend en moeizaam, herneming, doelwoord of -uiting onduidelijk, betekenis onduidelijk, expressieve afasie en algemene communicatiebeperking een uitstekende IBB vertoonden.

3.3 Concurrente validiteit van APROCSA-NL

De correlatiematrix met Benjamini-Hochberg gecorrigeerde Spearman correlaties tussen de APROCSA-NL features en de kwantitatieve natuurlijke spraakfeatures enerzijds en testscores anderzijds zijn weergegeven in Figuur 4.

Op voorhand werden verschillende APROCSA-NL features gelinkt aan deze kwantitatieve natuurlijke spraakfeatures en testscores. Deze 59 paren zijn aangeduid met een zwarte omkadering in Figuur 4. Van deze 59 paren, waren 51 significant met elkaar gecorreleerd, waarvan er 30 sterk ($|r| \geq 0.50$) met elkaar gecorreleerd waren. Daarnaast konden we 23 van de 28 APROCSA-NL features linken aan minstens één kwantitatieve natuurlijke spraakmaat of testscore om de concurrente validiteit na te gaan. Van deze 23 APROCSA-NL features, waren er 19 significant gecorreleerd met hun kwantitatieve tegenhanger, waarvan er 14 sterk ($|r| \geq 0.50$) met elkaar gecorreleerd waren. We vonden geen significante verbanden met de kwantitatieve tegenhangers voor de volgende APROCSA-NL features: neologismen, omissie van het gebonden morfeem, herneming en doelwoord of -uiting onduidelijk.

Figuur 4: Spearman correlaties tussen de APROCSA-NL features en de kwantitatieve natuurlijke spraakmaten en testcores



Enkel significante correlaties (Benjamini-Hochberg gecorrigeerde $p < 0.05$) worden weergegeven. Negatieve correlaties zijn aangeduid in het blauw, positieve correlaties in het rood. In het zwart omkaderd werden de a priori gedefinieerde relaties tussen APROCSA-NL features en hun kwantitatieve tegenhanger aangeduid.

3.4 Onderliggende constructen van APROCSA-NL

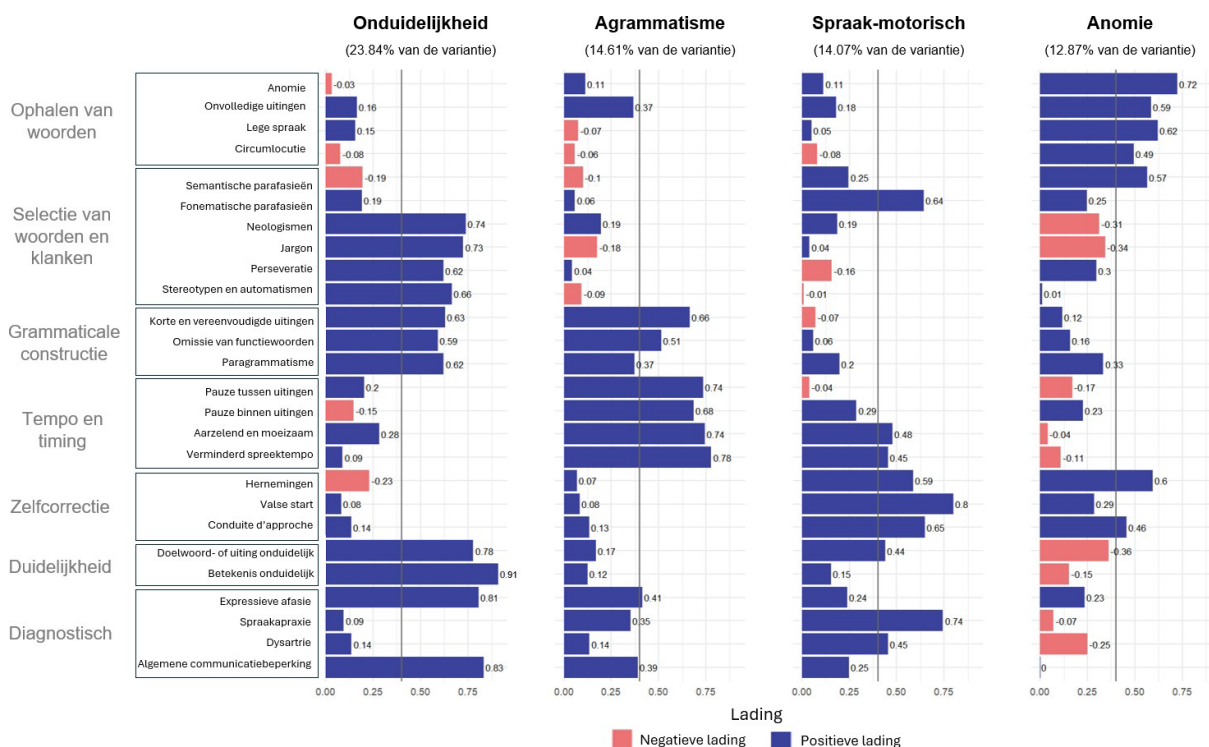
We voerden een factoranalyse met varimax rotatie uit om de onderliggende constructen van APROCSA-NL te onderzoeken bij personen met afasie na een beroerte. Van de initiële 28 APROCSA-NL features, werden er twee geëxcludeerd (ommissie van het gebonden morfeem, off topic) aangezien deze een lading < 0.4 op alle factoren of een uniciteit > 0.75 hadden, wat betekent dat ze niet betekenisvol bijdragen aan de factorstructuur. Op basis van de overblijvende 26 features vonden we evidentie voor vier onderliggende constructen, die 65.4% van de totale variantie in natuurlijke spraakfeatures verklaarden. De factorladingen worden weergegeven in Figuur 5.

De eerste factor (Onduidelijkheid) verklaarde 23.84% van de totale variantie en reflecteerde voornamelijk moeilijkheden met grammaticale constructies en het selecteren van woorden en klanken (neologismen, jargon, perseveraties en stereotypen en automatismen). Dit resulteerde in expressieve afasie en een algemeen communicatieprobleem, wat de duidelijkheid van de natuurlijke spraak sterk aantastte.

De tweede factor (Agrammatisme) verklaarde 14.61% van de totale variantie en werd gekarakteriseerd door hoge ladingen op grammaticale constructies, met een invloed op het tempo en de timing van de natuurlijke spraak.

De derde factor (Spraak-motorisch) verklaarde 14.07% van de totale variantie en weerspiegelde een groep met comorbide dysartrie of spraakapraxie, resulterend in een atypische tempo en timing, zelfcorrecties en verminderde

Figuur 5: Factorladingen voor de APROCSA-NL features, gegroepeerd per factor



Positieve ladingen zijn weergegeven in het blauw, negatieve ladingen zijn weergegeven in het rood. De grijze verticale lijn duidt de lading cutoff van 0.4 aan.

duidelijkheid van de natuurlijke spraak. Belangrijk om op te merken is dat de hoge lading van fonematische parafasieën vermoedelijk eerder spraakapraxie dan fonologische afasie weerspiegelt. Deze hypothese wordt verder ondersteund door het feit dat deelnemers met comorbide spraakapraxie significant hogere Spraak-motorische factorscores hebben in vergelijking met individuen zonder comorbide spraakapraxie (Mann-Whitney $U = 80.00$, $p < 0.001$).

De vierde factor (Anomie) verklaarde 12.87% van de variantie in natuurlijke spraakfeatures en was veelal gerelateerd aan het ophalen van woorden en semantische parafasieën en resulteerde in zelfcorrecties.

4. Discussie

Dit artikel bespreekt APROCSA-NL, de Nederlandstalige bewerking van APROCSA (Casilio e.a., 2019). APROCSA werd vertaald van het Engels naar het Nederlands via *forward and*

backward translation. De meeste APROCSA-NL features vertoonden een goede tot uitstekende interbeoordelaarsbetrouwbaarheid voor zowel ervaren als onervaren beoordelaars, en waren valide metingen van natuurlijke spraak

We vonden vier onderliggende constructen van natuurlijke spraak: Onduidelijkheid, Agrammatisme, Spraak-motorisch en Anomie. Daarnaast toonden we aan dat APROCSA-NL een tijdsefficiënte aanpak biedt om natuurlijke spraak in kaart te brengen, aangezien de gemiddelde scoringsduur voor zowel ervaren als onervaren beoordelaars minder dan vijftien minuten bedroeg.

We vonden eveneens dat APROCSA-NL ook gebruikt kan worden bij korte natuurlijke spraaksamples van minder dan drie minuten of bij zeer ernstige afasie, met als belangrijke kanttekening dat niet alle APROCSA-NL features even betrouwbaar zijn voor alle samplelengtes en ernstgraden van afasie.

4.1 APROCSA-NL features

Zowel APROCSA als APROCSA-NL bevatten 27 natuurlijke spraakfeatures. Er zijn twee belangrijke verschillen tussen APROCSA en APROCSA-NL met betrekking tot de geïncloseerde features. In de originele APROCSA werd circumlocutie geëxcludeerd wegens een slechte interbeoordelaarsbetrouwbaarheid, terwijl deze feature in onze studie een goede ICC-waarde ($r = 0.68$) behaalde en dus wel behouden bleef. Daarentegen had omissie van het gebonden morfeem een zeer lage ICC ($r = 0.08$) in de Nederlandstalige studie, terwijl deze feature het uitstekend deed ($r = 0.83$) in de Engelstalige studie. Het verschil in morfologische structuur en frequentie van gebonden morfemen in het Engels en het Nederlands zou betrouwbare scoring van het feature ‘omissie van het gebonden morfeem’ kunnen verhinderen voor het Nederlands, wat de zeer lage ICC in de huidige studie zou kunnen verklaren.

4.2 Interbeoordelaarsbetrouwbaarheid van APROCSA-NL

In het algemeen is de IBB van de APROCSA-NL ($r = 0.68$) relatief lager in vergelijking met de IBB van de originele APROCSA ($r = 0.80$). Dit zou te maken kunnen hebben met het feit dat Casilio en collega's een diverse groep van personen met afasie selecteerden, wat het eenvoudiger maakte om hun verschillende natuurlijke spraakpatronen te identificeren. In deze studie daarentegen waren de deelnemers niet vooraf geselecteerd. Het feature ‘omissie van het gebonden morfeem’ had een opmerkelijk lagere ICC in de huidige studie en werd bijgevolg geëxcludeerd voor de Nederlandstalige versie.

Een recente studie vond dat ervaren logopedisten hun betrouwbaarheid in het scoren van de APROCSA kunnen verbeteren met training, wat een veelbelovende bevinding is om de beoordeling door één logopedist te rechtvaardigen (Casilio e.a., in press). Daarnaast suggereert de huidige studie dat - wanneer adequate training voorzien wordt - onervaren beoordelaars de APROCSA-NL met een gelijkaardige IBB scoren als ervaren beoordelaars.

In tegenstelling tot de originele APROCSA-studie, hanteerden we geen minimumcriterium van vijf minuten voor de natuurlijke spraaksamples. We wilden namelijk een heterogene groep van personen met afasie onderzoeken, waardoor het niet realistisch is om spraaksamples van vijf minuten te vereisen. We gingen bijkomend de invloed van zowel sampleduur als afasie-ernst op de IBB na.

Zelfs bij relatief korte natuurlijke spraaksamples van minder dan drie minuten, bereikten de meerderheid van de APROCSA-NL features een goede tot uitstekende IBB, wat het gebruik van APROCSA-NL ondersteunt in het geval dat er geen sample van vijf minuten verzameld kon worden. Deze resultaten zijn in overeenstemming met een latere studie van Casilio en collega's (2025), waaruit bleek dat natuurlijke spraaksamples van drie minuten nog steeds resulteerden in een goede IBB. Anderzijds toonden enkele features (bv. pauzes) een relatief lagere ICC voor langere ($>$ vijf minuten) in vergelijking met kortere (\leq vijf minuten) samples, wat zou kunnen suggereren dat deze features suboptimaal zijn voor auditief-perceptuele beoordeling, en best in kaart gebracht worden via een objectieve, automatische analyse. Wat de afasie-ernst betreft, is het veelbelovend dat vijftien APROCSA-NL features nog steeds een goede tot uitstekende IBB behaalden bij ernstige afasie. Samengevat, tonen de APROCSA-NL features een verschillend patroon van interbeoordelaarsbetrouwbaarheid afhankelijk van de duur van het natuurlijke spraaksample en de ernst van de afasie, wat belangrijk is om in het achterhoofd te houden wanneer de APROCSA-NL afgenomen en gescoord wordt.

De IBB van APROCSA-NL vergelijken met de IBB van andere Nederlandstalige natuurlijke spraakassessments is uitdagend door de verschillen in de geïncloseerde natuurlijke spraakfeatures en hun operationalisatie, maar ook door variatie in de deelnemers- en beoordelaarsgroep. Voor de schaal begripbaarheid van de ANTAT varieert de ICC van 0.92 tot 0.94, afhankelijk van de gebruikte ANTAT-versie (Blomert e.a., 1994), wat in overeenstemming is

met onze ICC's voor algemene communicatiebeperking ($r = 0.90$) en betekenis onduidelijk ($r = 0.86$).

De IBB voor de Nederlandstalige Main Concept Analysis werd berekend voor 35 personen met afasie en resulteerde in een ICC van 0.895 voor de effectiviteit van de informatieoverdracht en 0.933 voor de efficiëntie van de informatieoverdracht (Criel e.a., 2025).

De ASTA, wat een transcriptie-gebaseerd instrument is, maar inhoudelijk het dichtste aanleunt bij de APROCSA-NL features, behaalt een gemiddelde ICC van 0.79 voor negen features gerelateerd aan lexicale en morfosyntactische vaardigheden (Van Der Scheer e.a., 2011). Deze ICC's werden echter berekend voor vijftien neurologisch gezonde deelnemers, gescoord door vier linguïsten.

4.3 Concurrente validiteit van APROCSA-NL

Onze validiteitsanalyse includeerde 23 APROCSA-NL features die we a priori gelinkt hadden aan een transcriptie-gebaseerde natuurlijke spraakmaat of een testscore. Voor de meerderheid van deze APROCSA-NL features vonden we een significante correlatie met hun kwantitatieve tegenhanger, wat suggereert dat APROCSA-NL een valide instrument is om natuurlijke spraak bij afasie in kaart te brengen. We vonden geen significante correlaties voor de volgende APROCSA-NL features: neologisme, omissie van het gebonden morfeem, herneming, en doelwoord of -uiting onduidelijk. Herneming, en doelwoord of -uiting onduidelijk toonden initieel significante correlaties met hun kwantitatieve tegenhanger, maar deze overleefden de correctie voor meerdere vergelijkingen niet. Bijkomend waren er slechts tien deelnemers die zich presenteerden met neologismen of omissie van het gebonden morfeem, wat hun detectie in de validiteitsanalyse mogelijk verhinderd heeft.

4.4 Onderliggende constructen van APROCSA-NL

We vonden evidentie voor vier onderliggende constructen van natuurlijke spraak bij afasie gebaseerd op de APROCSA-NL-scores: Onduidelijkheid, Agrammatisme, Spraak-motorisch en Anomie. Drie van deze vier constructen lijken sterk op de constructen die gevonden werden in de originele APROCSA-studie. Zowel APROCSA als APROCSA-NL vonden namelijk evidentie voor een Agrammatisme ('Agrammatism') en Spraak-motorische ('Motor speech') factor.

Daarnaast is onze factor Anomie gelijkaardig aan de factor 'Logopenia' van Casilio en collega's, met dat verschil dat onze factor Anomie hoog laadt op alle features binnen de categorie 'ophalen van woorden' en semantische parafasieën. Bijkomend had onze factor Anomie geen hoge ladingen op perseveraties en 'tempo en timing' features, wat wel het geval was voor de factor 'Logopenia' uit de originele studie.

Tot slot vonden we een vierde factor, Onduidelijkheid. Deze factor werd geassocieerd met grammaticale constructie features, maar ook met neologismen, jargon, perseveraties en automatismen, die een sterke invloed hebben op de duidelijkheid van de natuurlijke spraak - vandaar 'Onduidelijkheid' - resulterend in ernstige manifestaties van expressieve afasie. De vierde factor uit de originele APROCSA-studie ('Parafasie') daarentegen reflecteerde het verkeerd selecteren van woorden en klanken en was geassocieerd met onder andere semantische parafasieën, fonematische parafasieën, paragrammatismen en valse starts. Deze features waren verspreid over verschillende constructen in de huidige studie (semantische parafasieën bij Anomie, fonematische parafasieën en valse starts bij Spraak-motorisch, paragrammatisme bij Onduidelijkheid). De factoren Agrammatisme en Spraak-motorisch lijken bovendien robuuste onderliggende constructen van natuurlijke spraak, aangezien eerdere studies van natuurlijke spraak (bv., Ding e.a., 2020; Gordon, 2020; Vermeulen e.a., 1989) deze constructen eveneens vonden.

4.5 Feedback van logopedisten op APROCSA-NL

We vroegen via een informeel gesprek feedback aan de ervaren beoordelaars over de scoringsprocedures, features en haalbaarheid in de klinische praktijk van de APROCSA-NL.

Het gebruik van het woord ‘indruk’ in de beschrijving van het feature Anomie (d.i., algemene indruk van woordvindingsproblemen) werd waargenomen als dubbelzinnig, aangezien de instructies van APROCSA-NL expliciet beschrijven om zo objectief mogelijk te scoren. Dit zou ervoor gezorgd kunnen hebben dat er variatie was in hoe beoordelaars het feature anomie gescoord hebben, wat op zijn beurt kan verklaren waarom de IBB en validiteit van dit feature eerder aan de lage kant waren, terwijl dit toch één van de meest voorkomende afasiekenmerken is.

Een andere opmerking had betrekking op het woord ‘pervasief’ in de beschrijving van de scoreschaal. Score 2 komt overeen met ‘frequent aanwezig, maar niet pervasief’, terwijl een score 3 overeenkomt met ‘matig ernstig, pervasief’. Om meer duidelijkheid te scheppen omtrent het woord ‘pervasief’, werd het volgende toegevoegd aan de handleiding: “Pervasief verwijst naar de impact of invloed van een feature op het volledige natuurlijke spraaksample. Score 3 (gemarkeerd) betekent dus dat de aanwezigheid van dit feature een significante impact heeft op het spraaksample.”

Ten slotte verwachtten de logopedisten dat het scoren van de APROCSA-NL haalbaar moet zijn in zowel acute als revalidatiesettings. Desalniettemin gaven ze aan dat het handig zou zijn om een overzicht van alle features en hun definitie te hebben op één pagina, om naast je te hebben tijdens het scoren. Dit is vooral interessant voor features die een meer gedetailleerde beschrijving hebben wat score 1, 2, 3 of 4 inhoudt (bv. bij de features korte en vereenvoudigde uitingen, omissie van functiewoorden of algemene communicatiebeperking). Om aan deze nood tegemoet te komen, creëerden we een APROCSA-NL ‘cheat sheet’ waar in één oog-

opslag alle features en hun definities en scorecriteria geraadpleegd kunnen worden.

Een laatste aanbeveling van de logopedisten was om video-opnames te gebruiken in plaats van enkel audio (zoals aanbevolen in de originele APROCSA-studie), aangezien dit helpend zou zijn om bijvoorbeeld zoekgedrag bij spraakapraxie te identificeren. Gegeven dat fonematische parafasieën sterk laadden op de factor Spraak-motorisch, is het niet ondenkbaar dat deze feature eerder spraakpractische fouten dan fonematische parafasieën weerspiegelde.

4.6 Limitaties en vervolgonderzoek

Deze studie heeft enkele limitaties. Ten eerste blijft de auditief-perceptuele aanpak inherent subjectief, waardoor deze vatbaar is voor variatie in de scoring afhankelijk van de ervaring van de beoordelaar. We probeerden deze invloed echter zoveel mogelijk te minimaliseren door voor elke beoordelaar dezelfde training te voorzien.

Ten tweede includeerden we in deze studie enkel personen met afasie ten gevolge van een beoorte. Hoewel de APROCSA-features destijds zo geselecteerd werden door Casilio en collega’s dat deze verscheidene kenmerken van afasie ten gevolge van diverse etiologieën weerspiegelen, is het aangeraden dat toekomstig onderzoek nagaat of APROCSA(-NL) ook een betrouwbaar en valide instrument is bij bijvoorbeeld primair progressieve afasie.

Ten derde droeg deze studie eveneens bij aan de verdere validatie van de APROCSA-methode door de invloed van sampleduur en afasie-ernst op de interbeoordelaarsbetrouwbaarheid na te gaan. We moeten hierbij echter opmerken dat we ongebalanceerde groepen hebben om de ICC’s te vergelijken tussen deze subgroepen van sampleduren en afasie-ernstgraden, wat zou kunnen leiden tot een over- of onderschatting van de ICC’s.

Ten vierde beschikten we in onze validiteitsanalyse niet over kwantitatieve tegenhangers voor vijf APROCSA-NL features: jargon, perseveraties, stereotypen en automatismen, betekenis

onduidelijk en off topic. Het includeren van macrostructurele kwantitatieve natuurlijke spraakmaten (bv. coherentie) is aangeraden in vervolgonderzoek om features zoals betekenis onduidelijk en off topic te valideren.

Ten vijfde hadden we een eerder klein sample van 58 personen met afasie om een factoranalyse met 26 features uit te voeren.

Ten slotte beschikten we niet over video-opnames. Toekomstig onderzoek zou kunnen onderzoeken of de betrouwbaarheid en validiteit van APROCSA-NL nog verder kan toenemen indien men zowel over audio- als video-opnames beschikt.

5. Besluit

Deze studie stelt APROCSA-NL voor, de Nederlandstalige bewerking van APROCSA (Casilio e.a., 2019). De meerderheid van de APROCSA-NL features waren valide metingen van natuurlijke spraak en behaalden een goede tot uitstekende interbeoordelaarsbetrouwbaarheid. We toonden met dit onderzoek eveneens aan dat APROCSA-NL gebruikt kan worden bij kortere samples van minder dan drie minuten en in gevallen van ernstige afasie. Bovendien bleek APROCSA-NL een tijdsefficiënte manier om natuurlijke spraak bij afasie in kaart te brengen, met een gemiddelde scoretijd van minder dan tien minuten.

Dankwoord

Bedankt aan alle logopedisten en personen met afasie die deelgenomen hebben aan deze studie. We hopen dat jullie APROCSA-NL een zinvolle aanvulling vinden binnen het afasie assessment. Daarnaast een welgemeende dankjewel aan Jelle Michiels voor haar hulp bij de *backward translation*, onze student-beoordelaars Irimi Beke, Frea Diddens en Anouk Verheyen en tot slot onze ervaren beoordelaars, Sarah Geens en Karen Peeters.

Financiering

Deze studie werd gefinancierd door het Fonds Wetenschappelijk Onderzoek (FWO) Vlaanderen, via een Aspirant Strategisch Basisonderzoek toegekend aan Mara Barberis (1SH1Q24N).

Contact

Mara Barberis
mara.barberis@kuleuven.be

Referenties

- Alfuraydan, M., Croxall, J., Hurt, L., Kerr, M., & Brophy, S. (2020). Use of telehealth for facilitating the diagnostic assessment of Autism Spectrum Disorder (ASD): A scoping review. *PLOS ONE*, *15*(7), e0236415. <https://doi.org/10.1371/journal.pone.0236415>
- Armstrong, L., Brady, M., Mackenzie, C., & Norrie, J. (2007). Transcription-less analysis of aphasic discourse: A clinician's dream or a possibility? *Aphasiology*, *21*(3-4), 355-374. <https://doi.org/10.1080/02687030600911310>
- Azios, J.H., Archer, B., & Lee, J.B. (2021). Detecting behavioural change in conversation: procedures and preliminary data. *Aphasiology*, *35*(7), 961-983. <https://doi.org/10.1080/02687038.2020.1812031>
- Beaton, D.E., Bombardier, C., Guillemin, F., & Ferraz, M.B. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine*, *25*(24), 3186-3191. <https://doi.org/10.1097/00007632-200012150-00014>
- Blomert, L., Kean, M.L., Koster, Ch., & Schokker, J. (1994). Amsterdam-Nijmegen everyday language test: Construction, reliability and validity. *Aphasiology*, *8*(4), 381-407. <https://doi.org/10.1080/02687039408248666>
- Boles, L., & Bombard, T. (1998). Conversational discourse analysis: Appropriate and useful sample sizes. *Aphasiology*, *12*(7/8), 547-560.
- Boxum, E., Van Der Scheer, F., Van Werven, N., Dirksen, W., & Jonkers, R. (2019). Toevoeging van nieuwe linguïstische en communicatieve maten binnen de huidige ASTA. In *Stem-, Spraak- en Taalpathologie* (Vol. 24, pp. 23-40). Secretariaat cluster Nederlands - Rijksuniversiteit Groningen. <https://doi.org/10.21827/5ce3b9f148e00>
- Boxum, E., van der Scheer, F., & Zwaga, M. (2013). *Analyse voor Spontane Taal bij Afasie*. Vereniging voor Klinische Linguïstiek.
- Boyle, M. (2014). Test-retest stability of word retrieval in aphasic discourse. *Journal of Speech, Language, and Hearing Research*, *57*(3), 966-978. https://doi.org/10.1044/2014_JSLHR-L-13-0171
- Boyle, M. (2015). Stability of word-retrieval errors with the Aphasiabank stimuli. *American Journal of Speech-Language Pathology*, *24*(4), S953-S960. https://doi.org/10.1044/2015_AJSLP-14-0152
- Brookshire, R.H., & Nicholas, L.E. (1994). Speech sample size and test-retest stability of connected speech measures for adults with aphasia. *Journal of Speech, Language, and Hearing Research*, *37*(2), 399-407. <https://doi.org/10.1044/jshr.3702.399>
- Bryant, L., Ferguson, A., & Spencer, E. (2016). Linguistic analysis of discourse in aphasia: A review of the literature. In *Clinical Linguistics and Phonetics* (Vol. 30, Issue 7, pp. 489-518). Taylor and Francis Ltd. <https://doi.org/10.3109/02699206.2016.1145740>
- Bryant, L., Ferguson, A., Valentine, M., & Spencer, E. (2019). Implementation of discourse analysis in aphasia: investigating the feasibility of a Knowledge-to-Action intervention. *Aphasiology*, *33*(1), 31- 57. <https://doi.org/10.1080/02687038.2018.1454886>
- Casilio, M., Dutta, M., Bryan, K., Crouch, K., Dewall, Z., Feld, I., Kile, J., McMahan, K., Samora, J., Shibata, K., Steel, S., Kasdan, A.V., Walljasper, L., de Riesthal, M., & Wilson, S.M. (in press). *A clinical training program for auditory-perceptual rating of connected speech in aphasia*.
- Casilio, M., Kasdan, A.V., Bryan, K., Shibata, K., Schneck, S.M., Levy, D. F., Entrup, J.L., Onuscheck, C., de Riesthal, M., & Wilson, S.M. (2025). Four dimensions of naturalistic language production in aphasia after stroke. *Brain*, *148*(1), 291-312. <https://doi.org/10.1093/brain/awae195>
- Casilio, M., Rising, K., Beeson, P.M., Bunton, K., & Wilson, S.M. (2019). Auditory-perceptual rating of connected speech in Aphasia. *American Journal of Speech-Language Pathology*, *28*(2), 550-568. https://doi.org/10.1044/2018_AJSLP-18-0192
- Cavanaugh, R., & Haley, K. L. (2020). Subjective communication difficulties in very mild aphasia. *American Journal of Speech-Language Pathology*, *29*(1S), 437-448. https://doi.org/10.1044/2019_AJSLP-CAC48-18-0222
- Cicchetti, D.V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, *6*(4), 284-290. <https://doi.org/10.1037/1040-3590.6.4.284>
- Criel, Y., De Groote, E., Benbassat, E., Beirinckx, B., Van Damme, M., Kong, A. P.-H., & De Letter, M. (2025). The main concept analysis: Sensitivity, specificity and inter-rater reliability in Dutch-speaking people with aphasia. *Journal of Communication Disorders*, *116*, 106543. <https://doi.org/10.1016/j.jcom-dis.2025.106543>
- Criel, Y., Deleu, M., De Groote, E., Bockstael, A., Kong, A.P.H., & De Letter, M. (2021). The dutch main concept analysis: Translation and establishment of normative data. *American Journal of Speech- Language Pathology*, *30*(4), 1750-1766. https://doi.org/10.1044/2021_AJSLP-20-00285
- Dietz, A., & Boyle, M. (2018). Discourse measurement in aphasia research: Have we reached the tipping point? In *Aphasiology* (Vol. 32, Issue 4, pp. 459-464). Routledge. <https://doi.org/10.1080/02687038.2017.1398803>
- Ding, J., Martin, R.C., Cris Hamilton, A., & Schnur, T.T. (2020). Dissociation between frontal and temporal-parietal contributions to connected speech in acute stroke. *Brain*, *143*(3), 862-876. <https://doi.org/10.1093/brain/awaa027>
- Fromm, D., Forbes, M., Holland, A., Dalton, S.G., Richardson, J., & MacWhinney, B. (2017). Discourse characteristics in aphasia beyond the Western Aphasia Battery cutoff. *American Journal of Speech- Language Pathology*, *26*(3), 762-768. https://doi.org/10.1044/2016_AJSLP-16-0071

- Gordon, J.K. (2020). Factor analysis of spontaneous speech in aphasia. *Journal of Speech, Language, and Hearing Research*, 63, 4127-4147. <https://doi.org/10.23641/asha>
- Hallowell, B. (2016). Aphasia and other acquired neurogenic language disorders: A guide for clinical excellence. In *ProtoView* (Vol. 3, Issue 11). Ringgold, Inc.
- Holland, A., Forbes, M., Fromm, D., & MacWhinney, B. (2019). Communicative strengths in severe aphasia: The Famous People Protocol and its value in planning treatment. *American Journal of Speech-Language Pathology*, 28(3), 1010-1018. https://doi.org/10.1044/2019_AJSLP-18-0283
- Kim, H., Berube, S., & Hillis, A.E. (2022). Core lexicon in aphasia: A longitudinal study. *Aphasiology*. <https://doi.org/10.1080/02687038.2022.2121598>
- Kim, H., & Wright, H.H. (2020). Concurrent validity and reliability of the core lexicon measure as a measure of word retrieval ability in aphasia narratives. *American Journal of Speech-Language Pathology*, 29(1), 101-110. https://doi.org/10.1044/2019_AJSLP-19-0063
- Le, D., Licata, K., & Mower Provost, E. (2018). Automatic quantitative analysis of spontaneous aphasic speech. *Speech Communication*, 100, 1-12. <https://doi.org/10.1016/j.specom.2018.04.001>
- Leaman, M.C., & Edmonds, L.A. (2021). Assessing language in unstructured conversation in people with aphasia: Methods, psychometric integrity, normative data, and comparison to a structured narrative task. <https://doi.org/10.23641/asha>
- Lesser, R., & Algar, L. (1995). Towards combining the cognitive neuropsychological and the pragmatic in aphasia therapy. *Neuropsychological Rehabilitation*, 5(1-2), 67-92. <https://doi.org/10.1080/09602019508520176>
- Lomas, J., Pickard, L., Bester, S., Elbard, H., Finlayson, A., & Zoghaib, C. (1989). The communicative effectiveness index: Developmental and psychometric evaluation of a functional communication measure for adult aphasia. *Journal of Speech and Hearing Disorders*, 54(1), 113-124. <https://doi.org/10.1044/jshd.5401.113>
- MacWhinney, B., Fromm, D., Forbes, M., & Holland, A. (2011). Aphasiabank: Methods for studying discourse. *Aphasiology*, 25(11), 1286-1307. <https://doi.org/10.1080/02687038.2011.589893>
- McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017). Montreal Forced Aligner: Trainable text-speech alignment using Kaldi. *InterSpeech 2017*, 498-502. <https://doi.org/10.21437/InterSpeech.2017-1386>
- Oostdijk, N. (2000). The Spoken Dutch Corpus: Overview and first evaluation. *Proceedings of LREC- 2000, Athens*, 2.
- Pakhomov, S.V.S., Smith, G.E., Chacon, D., Feliciano, Y., Graff-Radford, N., Caselli, R., & Knopman, D.S. (2010). Computerized analysis of speech and language to identify psycholinguistic correlates of frontotemporal lobar degeneration. *Cognitive and Behavioral Neurology*, 23(3), 165-177. <https://doi.org/10.1097/WNN.0b013e3181c5dde3>
- Parsa, M., Alam, M.R., & Mihailidis, A. (2021). *Towards AI-powered language assessment tools*. Poncelet, J., & Van hamme, H. (2022). Learning to jointly transcribe and subtitle for end-to-end spontaneous speech recognition. *Proceedings SLT 2022*. <https://doi.org/10.48550/arXiv.2210.07771>
- Pritchard, M., Hilari, K., Cocks, N., & Dipper, L. (2018). Psychometric properties of discourse measures in aphasia: acceptability, reliability, and validity. *International Journal of Language and Communication Disorders*, 53(6), 1078-1093. <https://doi.org/10.1111/1460-6984.12420>
- Ruiter, M.B., Otters, M.C., Piai, V., Lotgering, E.A.M., Theunissen, J.E.M.C., & Rietveld, T.C.M. (2022). A transcription-less quantitative analysis of aphasic discourse elicited with an adapted version of the Amsterdam-Nijmegen Everyday Language Test (ANELT). *Aphasiology*, 1-20. <https://doi.org/10.1080/02687038.2022.2109124>
- Sanborn, V., Ostrand, R., Ciesla, J., & Gunstad, J. (2020). Automated assessment of speech production and prediction of MCI in older adults. *Applied Neuropsychology: Adult*. <https://doi.org/10.1080/23279095.2020.1864733>
- Sousa, V.D., & Rojjanasrirat, W. (2011). Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: a clear and user-friendly guideline. *Journal of Evaluation in Clinical Practice*, 17(2), 268-274. <https://doi.org/10.1111/j.1365-2753.2010.01434.x>
- Stark, B.C. (2023). Historical review of research in discourse deficits and its recent advancement. In *Spoken Discourse Impairments in the Neurogenic Populations* (pp. 3-22). Springer International Publishing. https://doi.org/10.1007/978-3-031-45190-4_1
- Stark, B.C., & Dalton, S. G. (2024). A scoping review of transcription-less practices for analysis of aphasic discourse and implications for future research. *International Journal of Language & Communication Disorders*, 59(5), 1734-1762. <https://doi.org/10.1111/1460-6984.13028>
- Stark, B.C., Dutta, M., Murray, L.L., Fromm, D., Bryant, L., Harmon, T.G., Ramage, A.E., & Roberts, A.C. (2021). Spoken discourse assessment and analysis in aphasia: An international survey of current practices. *Journal of Speech, Language and Hearing Research (JSLHR)*, 64(11), 1-24. https://doi.org/10.1044/2021_JSLHR-20-00708
- Stark, B.C., & Fukuyama, J. (2021). Leveraging big data to understand the interaction of task and language during monologic spoken discourse in speakers with and without aphasia. *Language, Cognition and Neuroscience*, 36(5), 562-585. <https://doi.org/10.1080/23273798.2020.1862258>
- Tan, Z.-H., Sarkar, A.kr., & Dehak, N. (2020). rVAD: An unsupervised segment-based robust voice activity detection method. *Computer Speech & Language*, 59, 1-21. <https://doi.org/10.1016/j.csl.2019.06.005>

Van Der Scheer, F., Zwaga, M., & Jonkers, R. (2011). Normering van de ASTA, analyse voor spontane taal bij afasie. *Stem-, Spraak- En Taalpathologie*, 17(2), 19-30.

Van Ewijk, L., Dijkhuis, L., Hofs-van Kats, M., Hendrickx, M., Wijngaarden, M., & De Hilster, C. (2018). *Nederlandse benoemtest (NBT)*. Bohn Stafleu van Loghum.

Vandenborre, D., Cornette, C., & Visch-Brink, E. (2021). *CAT-NL White paper: Conceptuele Analyse Situatieplaat*.

Vandenborre, D., Visch-Brink, E., van Dun, K., Verhoeven, J., & Mariën, P. (2018). Oral and written picture description in individuals with aphasia. *International Journal of Language and Communication Disorders*, 53(2), 294-307. <https://doi.org/10.1111/1460-6984.12348>

Vermeulen, J., Bastiaanse, R., & Van Wageningen, B. (1989). Spontaneous speech in aphasia: A correlational study. *Brain and Language*, 36(2), 252-274.

Visch-Brink, E.G., van de Sandt-Koenderman, M., & El Hachioui, H. (2010). *ScreeLing*. Bohn Stafleu Van Loghum.

Visch-Brink, E., Vandenborre, D., de Smet, H.J., Mariën, P., Swinburn, K., Porter, G., & Howard, D. (2015). *Comprehensive Aphasia Test – Nederlandstalige bewerking (CAT-NL)*. Pearson.

Wallace, S.E., & Kimelman, M.D.Z. (2013). Generalization of word retrieval following semantic feature treatment. *NeuroRehabilitation*, 32, 899-913. <https://doi.org/10.3233/NRE-130914>

Wild, D., Grove, A., Martin, M., Eremenco, S., McElroy, S., Verjee-Lorenz, A., & Erikson, P. (2005). Principles of good practice for the translation and cultural adaptation process for Patient-Reported Outcomes (PRO) measures: Report of the ISPOR task force for translation and cultural adaptation. *Value in Health*, 8(2), 94-104. <https://doi.org/10.1111/j.1524-4733.2005.04054.x>