Záróbeszámoló a K115385 pályázatról

1. Adatfelvétel

Az adatfelvétel lezárult. A végső statisztikákat lásd 1. sz melléklet. A publikált cikkek adatait folyamatosan közzétesszük a kutatói közösség számára a Wordbank (http://wordbank.stanford.edu/) nyilvános adatbázisban.

2. Viselkedéses vizsgálatok

a) Korábban feltártuk, hogy a Covid járványhelyzet által érintett 18 hónapos kisgyermekeket kevesebb szemkontaktus-jellegű közös figyelem kezdeményezés jellemezte a Korai Társas Kommunikáció Skála teszthelyzetben. A további elemzésekbe bevontuk az időközben feldolgozásra került Bayley-III Fejlődésteszt Nyelvi Skála adatait, amely szintén egy strukturált, interakción alapuló eljárás. Míg a kérdőíves anyai beszámolóval (KOFA-I) nyert receptív szókincs változón a járvány alatt vizsgált csoport magasabb pontszámot ért el, addig a Bayley-III Nyelvi Skála sztenderd pontszámban nem találtunk szignifikáns csoportkülönbséget. A kisgyermek otthonán kívüli társas tapasztalatainak mennyiségét a kérdőívekből kinyert, az intézményi gondozásra (pl. bölcsőde) vonatkozó mutatóval jellemeztük, ami nem állt összefüggésben más vizsgált változókkal, így nem magyarázhatta a kapott eredményeket. A bővített eredményeket konferenciákon prezentáltuk.

Konferencia előadások és absztraktok:
Kas, B., Tóth, I., Kolcsár, Sz.R., Harmati-Pap, V. (2022). Social-communicative skills and language development at 18 months: differences related to the Covid-19 pandemic. 11th European Congress of Speech and Language Therapy, 26-28 May, 2022, Salzburg, Austria.
Tóth, I., Kas, B., Kolcsár, Sz.R., Harmati-Pap, V. (2022). Másfél évesek társas-kommunikatív készségei és szókincsfejlődése járványhelyzetben. Nyelvfejlődés csecsemőkortól kamaszkorig 2. Budapest, 2022 február 2., ELTE BTK, 2022.
Tóth, I., Kolcsár, Sz.R., Harmati-Pap, V., Kas, B. (2022). Links between social-communicative behaviour and language development in a sample of Hungarian 18-month-olds: what is different in the Covid-19 pandemic? BCCCD 2022 Budapest CEU Conference on Cognitive Development. Program and Abstracts, p. 217, Paper A-0156.

b) A járványhelyzet gyermeki kommunikációra gyakorolt hatásának magyarázatához a 18 hónapos csecsemők egy almintáján (N=50), keresztmetszeti elrendezésben vizsgáltuk az anyai gondozási minőség és a kérdőívvel (KOFA-I) felmért szókincs kapcsolatát. A tízperces, laboratóriumban rögzített anya-csecsemő interakciók szenzitivitás szempontú elemzésével kimutattuk, hogy a gyermek jelzéseire, igényeire megfelelően reagáló és a gyermekkel kooperáló anyai viselkedés, valamint a magasabb anyai iskolázottság pozitívan korrelált a kisgyermek 18 hónapos kori expresszív szókincsével. Emellett a járvány alatt vizsgált csoportban (N=21) jelentősen magasabb volt az anyai szenzitivitás átlagpontszám, mint a járvány előtti csoportban (N=29). A moderátor modellben a Covid-érintettség nem volt szignifikáns, tehát nem módosította a szenzitivitás és az expresszív szókincs összefüggését. A mediátor modellben a szenzitivitás és az iskolázottság együttesen a variancia 12,6%-át magyarázták, a szenzitivitásnak azonban nem volt szignifikáns mediáló hatása. Az eredmények arra mutatnak, hogy a járványhelyzet – bizonyos feltételek mellett – elősegíthette az anyák jobb ráhangolódását gyermekük igényeire, és ezzel párhuzamosan facilitálta a gyermek szóprodukciójának gyorsabb fejlődését (Komarovszky Zsófia, MA Szakdolgozat, ELTE PPK, 2022).

c) A temperamentum olyan biológiailag meghatározott jellemzők összessége, amelyek már kora gyermekkorban azonosíthatók. Ezek a temperamentumjellemzők hozzájárulnak az affektusszabályozáshoz, a figyelem irányításához és a motoros aktivitáshoz (Rothbart, 1981, 2007). Korábbi vizsgálatok azt találták, hogy a korai pozitív temperamentumvonások (pl. a csecsemők figyelmi kontrollja és az önszabályozás képessége) korrelálnak a nyelvelsajátítás hatékonyságával (pl. az első szavak megjelenésének ideje és a szókincsbővülés ideje és sebessége) az első két életév során. Longitudinális vizsgálatunk célja a longitudinális összefüggések vizsgálata volt a korai csecsemő temperamentuma és a későbbi nyelvi fejlődés között. A csecsemő temperamentumát 6, 9 és 18 hónapos korban szülői kérdőíves módszerrel vizsgáltuk, a nyelvi készségeket (receptív és expresszív szókincs) a MacArthur-Bates Kommunikatív Fejlődési Adattár I. magyar változatával mértük fel 18 hónapos korban. A temperamentumjellemzők közül a 9 és 18 hónapos korban mért Extraverzió (surgency) és a 9 hónapos korban mért Akaratlagos kontroll faktorok összefüggést mutattak a 18 hónapos korban mért receptív szókincs-pontszámokkal: a magasabb Extraverzió (surgency) és Akaratlagos kontroll pontszámmal rendelkező csecsemők jobb receptív szókincs-nyelvi megértési készséget mutattak. Az eredménye lehetséges magyarázata, hogy a magasabb Extraverziós (surgency) értékeket mutató csecsemők könnyebben léphetnek be kommunikációs helyzetekbe, és nagyobb elkötelezettséget mutatnak a felnőtt szociális partnerekkel szemben. Továbbá, a magasabb erőfeszítést igénylő kontrollt mutató csecsemők könnyebben irányíthatják és tarthatják fenn a figyelmüket egy kommunikációs helyzetben.

Konferencia előadások és absztraktok:
Balázs, A., Tóth, I. Kas, B, Lakatos, K (2023). Higher infant surgency related to better linguistic outcome at 18 month. BCCCD 2022 Budapest CEU Conference on Cognitive Development. Program and Abstracts, pp. 204-5, PC-023.

2. Anyai dajkanyelv lexikai, fonetikai és morfoszintaktikai jellemzői

a) A dajkanyelvi sajátosságok kommunikációban betöltött szerepének elemzése nagyban hozzájárulhat az akusztikai vizsgálatok eredményeinek értelmezéséhez, ezért az idei évben folytatódott a gyerekekhez szóló beszédben észlelhető érzelmi töltet vizsgálata. A dajkanyelv egyik fontos jellemzője az erős pozitív érzelmek kifejezése a baba iránt, amit számos akusztikai mérésen és érzelemmegítélésen alapuló elemzés megerősített. Vizsgálatunkban újszerű megközelítésben két dimenzió mentén tanulmányoztuk az anyai beszédben fellelhető érzelmek percepcióját: az érzelemerősségre (arousal) és az érzelmi töltetre (valence) adott pontszámok alapján a baba 0, 4, 8 és 18 hónapos korában. Az észlelt érzelmek minden korban erősebbek és pozitívabbak voltak, mint a felnőtthöz intézett beszédben, és 4 és 18 hónapos kor között nem mutattak lényeges változást. A 0 hónapos kori dajkanyelvet alacsonyabb pontszámok jellemezték, aminek egyik lehetséges oka a hiányzó beszélőalkalmazkodás az újszülöttek édesanyjának beszédében. Az észlelt érzelmek a beszédaktusok szerint eltérő mintázatot mutattak a dajkanyelvben és a felnőtthöz szóló beszédben. A kérés pozitívabban jelölt volt, mint a felkiáltás, az érzelem erőssége viszont hasonló volt a két beszédaktus esetén. A babákhoz szóló beszéd ugyanakkor mindkét beszédaktus esetén erősebb érzelmeket közvetített, mint a felnőtthöz intézett mondatok.

A szünetek alapvető percepciós kulcsai a beszéd egységekre (pl. tagmondatokra, frázisokra) tagolásának. Kimutatták, hogy a babák egészen fiatal korban az olyan beszédet preferálják, amelyben a szünetek egybeesnek a szintaktikailag meghatározható tagmondathatárokkal. A dajkanyelvi szünetek időtartamáról, gyakoriságáról és legfőképp a tagmondatokkal vagy más nyelvi egységekkel való viszonyáról kevés, ellentmondásos eredmény áll rendelkezésre. 14 édesanya meséjének spontán részét elemeztük longitudinálisan a baba 0, 6 és 18 hónapos korában. Az eredmények azt mutatták, hogy a szünetek nemcsak gyakoribbak dajkanyelvben, mint felnőttekhez szóló beszédben, hanem arányaiban még gyakrabban fordulnak elő tagmondathatáron, mint tagmondat belsejében. Továbbá a tagmondatok, valamint a szünetek közötti beszédszakaszok hossza szótagszámban mérve kevesebb volt a 18 hónapos babákhoz szóló beszédben, mint a felnőttekhez szólóéban. Az eredmények arra utalnak, hogy a felnőttek dajkanyelvben gyakrabban tagolnak szünetekkel és jobban jelölik a szintaktikai egységek, tagmondatok határait, ezzel segítve a gyerekek beszédmegértési folyamatát.

A beszédben az egységek határait nemcsak szünetekkel, hanem glottalizációval is jelölhetik, amely szintén segítheti a beszéd egységekre tagolását, és így a beszédmegértést is. A glottalizáció frázishatárjelölő szerepe a dajkanyelvben azonban szinte egyáltalán nem ismert. A vizsgálathoz 20 édesanya felvételeit elemeztük longitudinálisan (4, 8 18 hónapos korban). A frázishatárt megelőző -lak szótag magánhangzóján végeztünk akusztikai méréseket (H1*−H2*, CPP, HNR), és hasonlítottuk össze frázis belseji előfordulásával. Az eredmények azt mutatták, hogy a célszó utolsó magánhangzóját glottalizáltabban ejtették intonációs frázis határon, mint frázis belsejében mindkét regiszterben, tehát a beszédegségek mindkét regiszterben egyaránt jelöltek voltak. Az eredmények arra utaltak ugyanakkor, hogy a dajkanyelvben periodikusabb zöngeképzés volt jellemző a felnőttekhez szóló beszédhez képest. A későbbiekben percepción alapuló kategorizáció (leheletes, modális, glottalizált zönge) adhatja meg a mérőszámok által mutatott regiszterbeli és prozódiai határon lévő eltérések jelentőségét.

A szünetezéssel kapcsolatos eredményeket a Beszédtudomány - Speech Science folyóiratban tettük közzé. A leheletes zöngeminőség-vizsgálat első szakaszának eredményeit a Nyelvfejlődés csecsemőkortól kamaszkorig 2. konferencián, majd az átdolgozott verziót a 18th Conference on Laboratory Phonology (LabPhon18) konferencián adtuk elő, az angol nyelvű kézirat előkészületben. A dajkanyelvi glottalizációról szóló absztraktunkat elfogadták a Beszédkutatás – Speech Research 2023 konferenciára. Az észlelt dajkanyelvi érzelmek longitudinális összegzése az Általános Nyelvészeti Tanulmányok XXXIV c. kötetbenban jelent meg könyvrészletként, a beszédaktusok különböző típusainak érzelmi töltetéről az International Conference on Speech Prosody konferencián adtunk elő, amelynek konferenciakötetében jelent meg az angol nyelvű közlemény.

Publikációk:

Kohári, A., Deme, A., Reichel, D.U., Szalontai, Á., & Mády K. (2022). Tartalmas és funkciószavak időtartama csecsemőkhöz szóló beszédben. MANYE–Akadémiai Kiadó. Közlésre elfogadva.

Kohári, A., Harmati-Pap, V., & Mády, K. (2022). A dajkanyelv tagolódása 6 hónapos csecsemőkhöz szóló történetmesélésben. Beszédtudomány - Speech Science. Közlésre elfogadva.

Mády, K., Gyuris, B., Gärtner, H.-M., Kohári, A., Szalontai, Á., & Reichel, U. D. (2022). Perceived emotions in infant-directed narrative across time and speech acts. In Proceedings 11th International Conference on Speech Prosody, pp. 590–594.

Mády, K., Kohári, A., Szalontai, Á., & Uwe, D. R. (2022). Észlelt érzelemkifejezés a dajkanyelvben. In Általános nyelvészeti tanulmányok 34. Budapest: Akadémiai Kiadó, pp. 221–246.

Konferencia előadások és absztraktok:

Kohári, A., Garai, L., Reichel, U.D., & Mády, K. (2022). Voice quality modifications in Hungarian infant-directed speech. 18th Conference on Laboratory Phonology (LabPhon18). 2022. június 25., online.

Kohári, A., Garai, L., Reichel, U.D., & Mády, K. (2022). A leheletes zöngeminőség longitudinális vizsgálata a dajkanyelvben. Nyelvfejlődés csecsemőkortól kamaszkorig 2. 2022. február 2., online.

Kohári, A., Reichel, U.D., Szalontai, Á., & Mády, K. (2022). (elfogadott absztrakt). A frázishatár zöngeminőségének vizsgálata felnőttekhez és gyerekekhez szóló beszédben. Beszédkutatás – Speech Research 2023.

Mády, K., Gyuris, B., Gärtner, H.-M., Kohári, A., Szalontai, Á., & Reichel, U. D. (2022). Perceived emotions in infant-directed narrative across time and speech acts. 11th International Conference on Speech Prosody. 23-26 May 2022. Lisbon.

b) A csecsemőkhöz szóló beszédet (IDS) vagy dajkanyelvet sajátos akusztikai tulajdonságok és lexikai/ szintaktikai egyszerűsítések különböztetik meg a felnőttekhez szóló beszédtől (ADS). A lexikai és morfoszintaktikai jellemzőire azonban kevés vizsgálat irányult, különösen a magyar nyelvben. Korábbi tanulmányok ellentmondó következtetéseket vontak le arra vonatkozóan, hogy az IDS összetevői hogyan válnak összetettebbé válnak-e a gyermekek fejlettségének megfelelően, főképp arra nézve, hogy az anyai nyelvhasználat komplexitása lineáris vagy nem lineáris módon növekszik.

Első vizsgálatunkban a minimális komplexitást a csecsemők 6 hónapos kora körül feltételeztük. A mért jellemzők a beszédmennyiség (a megszólalások száma, szavak száma), a lexikai jellemzők (type-token arány, ige-névszó arány,) és a morfoszintaktikai jellemzők (a megnyilatkozás átlagos hossza, a mondatok száma) voltak, sztenderd képanyagon alapuló rögtönzött mesemondás során, longitudinális kutatásban. 22 anya-csecsemő pár vett részt a vizsgálatban a csecsemő 1 napos, 6 hónapos és 18 hónapos korában, melynek során az anyák egyrészt gyereküknek (IDS), másrészt egy felnőtt vizsgálatvezetőnek (ADS) mondták el a történetet. Mind az IDS-t, mind az ADS-t rögzítettük, a később lejegyzett szöveget morfológiai és lexikai gyakorisági elemzés során értékeltük. Az eredmények azt mutatják, hogy az IDS lexikai és morfoszintaktikai jellemzői a maximális egyszerűsítést a 18 hónaposoknál mutatják. A vizsgálatunkban szereplő legidősebb gyermekek esetében az anyák több mondattal mesélték a mesét, kevesebb szókincset alkalmazva több ismétléssel. Kifejezéseik szignifikánsan rövidebbek voltak. Összességében az anyai IDS kisebb egységekben nagyobb beszédbőség jellemzi. Eredményeink arra utalnak, hogy nem lineáris az anyai IDS-nek a gyermek életkorával való progressziója.

A következő vizsgálatunk fő célja volt, hogy korábbi elemzéseinket egy új, 24 hónapos kori mérési ponttal kiegészítve vizsgáljuk az anyai dajkanyelvi narratívák szintaktikai és lexikai jellemzőit, és feltárjuk, hogy vajon a 18 hónapos kori komplexitási szint a legegyszerűbb szakasz időszaka, vagy már a felszálló ág része az u-görbén. A jelen kutatásban 29 anya-gyermek pár 6, 18 és 24 hónapos korban felvett narratíváinak elemzését végeztük el. 29 anya-csecsemő pár vett részt a vizsgálatban a csecsemő 6, 18 és 24 hónapos korában, melynek során az anyák egyrészt gyereküknek (IDS), másrészt egy felnőtt vizsgálatvezetőnek (ADS) mondtak el spontán szövegezett történetet egy sztenderd képsorozat alapján. Mind az IDS-t, mind az ADS-t rögzítettük, a később lejegyzett szöveget morfológiai és lexikai gyakorisági elemzés során értékeltük. Statisztikailag szignifikáns eltérést találtunk a beszédbőségi mutatók esetén (szószám és megnyilatkozásszám), a szintaktikai szerkezetben (átlagos megnyilatkozáshossz) és a lexikai változatosság tekintetében (type-token arány, igei type-token arány) a 6, a 18 és a 24 hónapos IDS narratívák esetén, az azonos időben rögzített ADS narratívákkal összevetve. Az eltérések minden esetben markánsak ADS és IDS között, azonban a különböző életkorokban felvett dajkanyelvi narratívák között nem mutatható ki szignifikáns változás. Jelentős variancia volt megfigyelhető abban, hogy az anyák a gyerekük melyik életkori szakaszában módosítják jobban beszédüket és milyen irányban. Így azt az eredményt, hogy a teljes mintát tekintve nem találunk szignifikáns eltérést az életkorok között az IDS-mutatókban, a csoporton belüli beszédmódosítási tendenciák sokszínűsége, azok egymást kioltó hatása okozhatja.

Publikációk:

Harmati-Pap, V., Vadász, N., Kas, B. & Tóth, I. (2021). Anyai dajkanyelvi narratívák lexikai és szintaktikai jellemzőinek longitudinális vizsgálata. Beszédtudomány, 2(1), 207-242. https://ojs.mtak.hu/index.php/besztud/issue/view/672/414

Harmati-Pap, V., Vadász, N., Tóth, I. & Kas, B. (2022). A lexikai és szintaktikai adaptáció idői mintázata az anyai dajkanyelvben. In: Bóna, Judit; Murányi, Sarolta (szerk.) A nyelvfeJlődés folyamata koragyermekkortól kamaszkorig, pp. 9-28. Budapest: ELTE. https://www.eltereader.hu/media/2022/11/Bona-Muranyi-A-nyelvfejlodes-folyamata-web4.pdf

3. EEG vizsgálatok

A statisztikai tanulás lehetővé teszi a környezeti események valószínűségi viszonyainak pontos modellezését és ezen keresztül a percepciót és a viselkedést is befolyásoló előrejelzések, perdikciók létrehozását. Vizsgálatainkban arra kerestünk választ, hogy mik ezekek a folyamatoknak a peremfeltételei újszülöttek esetében. Sikerült kimutatnunk, hogy a statisztikai tanulás már újszülött korban is jelentősen kontextus függő és képes a konkrét gyakoriságon felüli tágasabb kontextust is integrálni a környezetről alkotott modellbe. A valószínűségi változások integrálódnak a modell felépülése során, azonban amikor ezek a változások már nem bírnak további információs értékekkel, a jelenlétük egy tágabb modell részévé válik. Eredményeinket cikk formájában publikáltuk.

A statisztikai tanulás fontos elemének tűnik a bemutatott események idői elrendezése is. Bár a felnőttek képesek pusztán az események sorrendje alapján perdikciós modelleket alkotni az elkövetkező eseményekről, vizsgálatunk alapján úgy tűnik, hogy az események sorrendje (átmeneti valószínűsége) nem elegendő gyermekek esetében és legalábbis hallási kontextusban szükség van egy előre jósolható időzítésre is. Az eredményeket bemutattuk nemzetközi konferencián és a cikkünk kézirata jelenleg elbírálás alatt van egy nemzetközi folyóiratnál.

A nyelvi feldolgozásban fontos szerepet játszó egyszerű idői viszonyokat vizsgáló kísérleteink lehetőséget biztosítottak az idői feldolgozás fejlődésének vizsgálatára. Eredményeink alapján jelentős fejlődési változások figyelhetők meg az idői változásokat tükröző agyi kiváltott válaszok morfológiájában és agyi forrásaikban is. Az viszonylag egyszerű paradigma arra is lehetőséget biztosított számunkra, hogy az elsők között alkalmazzuk a pontos életkori anatómiai modelleken alapuló EEG forráslokalizációt egy hallási fejlődéssel kapcsolatos vizsgálatban. Cikkünk kézirata jelenleg elbírálás alatt van egy nemzetközi folyóiratnál.

A funkcionális hálózatok elemzésének módszertanát kidolgoztuk és mindhárom mért életkotban (0, 4 és 9 hó) alkalmaztunk mind FB és Dny különbségek kimutatására. Az adatok alapján a nyelvi feldolgozásért felelős hálózatok az életkorral egyre integráltabbá (összekötöttebbé) válnal és a jobb félteke irányába tolódnak el. Feltehető továbbá, hogy a dajkanyelv-függő hálózati topológia változások a dajkanyelv első év során betöltött szerepének változásával magyarázhatók. A kéziratunkat a lektorálás folyamán felmerült új szempontok alapján átdolgozzuk és hamarosan újra benyújtjuk a kéziratot egy nemzetközi folyóirathoz

A nyelvfejlődés előrejelzésében külön vizsgáltunk két kiemelt változót. Megállapítottuk, hogy a familiarizálás során bemutatott álszavak szótagjai közötti kapcsolatok és az ismétlődő szótagmintázatoktól eltérő ingerekre adott válaszok erőssége kapcsolatban áll a későbbi beszédképességgel, illetve jelezheti a diszlexiára való hajlamot.

A nyelvfejlődét előrejelző változók összetett statisztikai elemzése elkészült. Az EEG funkcionális hálózatokról kapott és az egyszerű idői viszonyok észlelését vizsgáló kísérletekben nyert változók is szerepelnek a nyelvi fejlődést statisztikailag legjobban előrejelző 25 változó között. Az eredménynek alapján értelmezésünk szerint az agyi hálózatok nagyobb integrációja áll a statisztikai bejósló erő hátterében, de az idői viszonyok feldolgozásának is szerepe van. Az eredményekről a jelentéshez csatolt statisztikai modellezést bemutató beszámoló ad részletesebb képet, amely későbbi publikációk alapjául fog szolgálni.

Publikációk:

Háden, G. P., Bouwer, F. L., Honing, H., & Winkler, I. (2022, preprint). Beat processing in newborn infants cannot be explained by statistical learning based on transition probabilities. bioRxiv. doi: 10.1101/2022.12.20.521245

Háden, G. P., Tóth, B., & Winkler, I. (2023, preprint). Longitudinal study of functional brain networks for processing infant directed and adult directed speech during the first year. bioRxiv. doi: 10.1101/2023.01.25.525490

Polver, S., Toth, B., Háden, G.P., Bulf, H., & Winkler, I. (2023, preprint). Early maturation of sound duration processing in the infant brain. bioRxiv. doi: 10.1101/2023.01.24.525349

Suppanen, E., Winkler, I., Kujala, T., & Ylinen, S. (2022). More efficient formation of longer-term representations for word forms at birth can be linked to better language skills at 2 years. Developmental Cognitive Neuroscience, 55, 101113. doi: 10.1016/j.dcn.2022.101113

Todd, J., Háden, G. P., & Winkler, I. (2022). Relevance to the higher order structure may govern auditory statistical learning in neonates. Scientific reports, 12(1), 1-10. doi: 10.1093/cercor/bhab344

Virtala, P., Kujala, T., Partanen, E., Hamalainen, J.A., & Winkler, I. (2023). Neural phoneme discrimination in variable speech in newborns; associations with dyslexia risk and later language skills. bioRxiv. doi: 10.1101/2023.01.10.522962

Konferencia előadások és absztraktok:

Háden, G. P., Bouwer, F. L., Honing, H., & Winkler, I. (2022). Beat and sequence representation in newborn infants. In: BCCCD 2022 (Budapest CEU Conference on Cognitive Development): Program and Abstracts p. 257, A-0220.

Winkler, I., Todd, J., & Háden, G.P. (2022). Statistical learning and higher-order structure in neonates. In MMN 2022 (The 9th Mismatch Negativity Conference, Fukushima, Japan), p. 93, Talk 2.

4. Gépi tanulási modell felállítása

A 18. hónapban mért receptív és expresszív szókincs előrejelzésére gépi tanulási modelleken alapuló elemzést végeztünk több mint kétezer egyedi lehetséges prediktor bevonásával. A modellezési módszertan az ún. Random Forest gépi tanulási algoritmusra épült. A modellezésből levonható következtetések általánosíthatósága érdekében kereszt-validálási eljárást alkalmaztunk, így a teljes elemzési láncra (imputálás, változószelekció, végső modell tanítása) érvényesült, hogy az adott modell előrejelzésének tárgyát képező minta semmilyen formában nem képezte részét a modell kialakításához és tanításához vezető lépéseknek. (A módszer pontos leírását az 1. sz. melléklet tartalmazza).

Azt találtuk, hogy az olyan korai mutatók, mint az anya életkora, BIG5 skálán mért extraverziója, a csecsemővel való kommunikációjának bizonyos fonetikai jellegzetességei és a csecsemő korban felvett EEG vizsgálatok bizonyos mutatói, noha mérsékelt pontossággal, de képesek előre jelezni a gyermek szülői beszámolókon alapuló 18. havi receptív szókincsét.

Konferencia előadások és absztraktok:

Háden, G.P. Tóth, B., Tóth, I., Tóth, D., Kohári, A., Kas, B., Mády, K., & Winkler, I. (2023). Neuro-cognitive predictors of early language development, In: BCCCD 2023 (Budapest CEU Conference on Cognitive Development): Program and Abstracts p. 198, PC-016.

5. Összegzés

Összességében, bár az adatok a COVID járvány hatására az eredeti tervtől eltérően nem egyetlen longitudinális mintaként kerültek felvételre, és egyes eredmények, köztük a nyelvelsajátítás előrejelzésre készített modell (1. sz. melléklet) publikálására a pályázati időszakon túl kerül sor, a projekt, a kitűzött célokat döntő mértékben teljesítette.

1. sz. melléklet: Report on predicting infants' language quality at 18 months from demographic, medical, neural, and parental speech quality measures collected before 9 months of age.

Report on
Predicting Infants' Language Quality at 18 Months from Demographic, Medical, Neural, and Parental Speech Quality Measures Collected before 9 Months of Age

Introduction

Predicting infants' quality of language acquisition is important for determining the crucial abilities, cognitive and neural processes supporting language acquisition. It has also widespread prospective societal benefits because it could allow early detection of potential developmental issues as well as early interventions, which are typically more effective due to the higher plasticity of the developing brain.

In the current prospective study, we collected a wide range of data about the medical status, neural and cognitive capabilities of the infants and the demographics, personality, and speech characteristics of the mother, in order to predict the quality of the infants speech comprehension and production at 18 months of age. Infants' speech quality was assessed with standardized tool. These outcome measures, some of the neural responses, and the mother's speech characteristics were repeatedly assessed to evaluate their change during the infant's early development.

The data were then entered into a statistical model based on machine learning, which tested how well the outcome measures could be predicted from the measures taken at an earlier age, and which measures proved to be most important for predicting speech quality at 18 months.

In the current report, we describe the methods of the data collection and statistical analysis, the main results and an assessment of the most important variables supporting the prediction of the outcome measures.

Methods

Recruitment, participants

Healthy, full-term infants were recruited 1-3 days after birth at the maternity ward of the Military Hospital Medical Centre in Budapest (N=75). The sample size was increased by additional recruitment through the social media, and eligible families joined the study at the 4-month data collection point (N=42). Native Hungarian speaking mothers and their first-born infants with birthweight ≥ 2500 grams and/or ≥ 37 gestational age were enrolled. Attrition throughout the study and temporary suspension of laboratory work due to the Covid-19 pandemic resulted in different sample sizes for each data collection point and measure.

Demographic and medical data

Mothers completed a demographic questionnaire about the infant and family at the first data collection point. Vast majority of the mothers (98.3%) were living with a spouse or partner, and 88.8% of infants were born after a planned pregnancy. As it is common in developmental studies with voluntary participants, mothers with higher education were overrepresented, 71.8% having college or university degree. Variables describing further characteristics of the sample are presented in Table 1.

Table 1. Sample characteristics (N=117).

Categorical and ordinal variables	Frequency (%)
Infant sex	
Male	56 (47.9)
Female	61 (52.1)
Maternal education	
Primary school	1 (.9)
Vocational school	4 (3.4)
Secondary school	15 (12.9)

Postsecondary applied education	13 (11.1)
College degree (BA, BSc)	35 (29.9)
University degree (MA, MSc, MD)	49 (41.9)
Number of people living in the same household*	
3	100 (86.2)
4 to 6	16 (13.8)
Regular financial problems in the family*	
None	114 (98.3)
Yes	2 (1.7)
Continuous variables	**Mean (SD); range**
Infant birthweight (grams)*	3371 (385); 2350–4120
Gestation at birth (weeks)	39.45 (1.18); 36–42
Maternal age at birth (years)	31.05 (4.81); 22–44
Paternal age at birth (years)*	33.72 (5.58); 23–52

*: N=116

Medical data were collected on maternal weight before pregnancy and after delivery, smoking habits, number of ultrasound scans during pregnancy, and method of delivery (Caesarean / normal).

Variables with appropriate variability were used in the statistical models to test the possible effects of socio-economic and birth circumstances on the 18-month language measures.

Language outcome measures

Infants' language outcomes were measured using two different methods: (1) structured parent report and (2) a standardized clinical test of expressive and receptive communication:

(1) Parent report. The Hungarian adaptation of the MacArthur-Bates CDI - Infant form (Words and Gestures); [21, 139]) was administered at 9 months and then bimonthly at 12, 14, 16 and 18 months of age in online or printed format according to the parent's preference.

(2) Clinical test of expressive and receptive communication. The Bayley Scales of Infant and Toddler Development, Third Edition, Language Scale: Expressive and Receptive Communication Subtests was administered in the lab at 9 and 18 months by trained experimenters.

Description and scoring of the CDI-I

The CDI obtains information about the language development of the children through the pre-structured, written report of their parents. The CDI-I: Words and Gestures Forms are for children ages 8–18 months in typical development. In the form, the first part prompts parents to document the child's understanding of hundreds of early vocabulary items separated into semantic categories such as animal names, household items, and action words. Parents mark the words understood or used, and the forms yield separate indexes of words understood and words produced. The second part of each form asks parents to record the communicative and symbolic gestures the child has tried or completed.

Four main variables has been calculated to represent each infant's language outcomes at each of the five time points (9, 12, 14, 16 and 18 months): (i) the number of different words the child understands (Receptive vocabulary), (ii) the number of different words the child says (Expressive vocabulary), (iii) the number of communicative and symbolic gestures the child uses (Gestures), (iv) the number of example sentences the child understands (Sentences).

Measuring maternal personality

Mothers filled in the 44-item Big5 questionnaire on the five basic personality traits (Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness to experience) as part of the 6.5-months data collection wave (N=108). Each scale had acceptable reliability (Cronbach alphas > .70) and high variability.

Measurement of maternal speech

Mothers were asked to perform the common everyday task of telling a story to their baby/infant using a booklet full of pictures. The booklet also contained certain fixed sentences that speakers were assumed to incorporate in their otherwise spontaneously told story based on the pictures. Recordings of the 17 fixed sentences formed the material for our analysis of acoustic parameters. The experimental sessions were performed as follows: first, mothers familiarized themselves with the book, then they told the story to the experimenter (AD condition), and finally to their infant (ID condition). Thus, all fixed sentences were spoken both under AD and ID conditions, enabling us to contrast the two registers in terms of acoustic parameters. Participants were recruited at the Military Hospital Medical Centre in Budapest in Budapest, where the recording sessions were conducted in a quiet room a few days after childbirth. The experiments were repeated at the age of 6 months of the infant in the baby lab of the Institute of Cognitive Neuroscience and Psychology, Research Centre for Natural Sciences. Speech samples at both locations were recorded using a Beyerdynamic TG H74c supercardioid head-mounted condenser microphone, digitized at a sampling rate of 44.1 kHz and a resolution of 16 bits.

For the acoustic analyses, the target utterances were first annotated in Praat. Fundamental frequency (f0) and vowel formants were calculated using the Praat software (version 6.0.37) (Boersma, 2001). In total, 196 phonetic parameters were extracted from each mother's recordings, using the toolkit CoPaSul (Reichel 2021). We investigated speech characteristics which have been associated with infant-directed speech in previous research (Saint-Georges 2013, Hilton et al. 2022). Thus, we analyzed variables related to fundamental frequency, speech rate, formant frequencies, speech intensity, and speech rhythm (for a complete list of the variables and the applied methodology, see Reichel 2021). F0-related parameters were transformed to semitones in order to account for interspeaker differences. Intensity was measured in terms of root mean squares (window length 50 ms, stepsize 10 ms). Signal and text were aligned using the BAS WebMAUS service (Kisler 2017). The values obtained for the fixed sentences were averaged for each mother separately for the two time points (0 and 6 months of the baby) and for both registers (IDS and ADS), yielding four means for a speaker. We compared each mother's IDS-means, and also the difference of their IDS- and ADS-means with the baby's receptive vocabulary at the age of 18 months. Furthermore, based on the extracted formants and the signal-text alignment we calculated the formants and formant space of all vowels in the entire story spoken by the mothers. Analogously to Bradlow et al. (1996) we define the formant space as the mean Euclidean distance of all vowels' formants F1 and F2 to the F1-F2 centroid.

EEG measurements

EEG was collected at 0, 4, and 9 months in sessions combining multiple EEG experiments. Variables from there of these experiments three (**Network study**, **Timing study**, **Resting State Network**) were included in the final statistical model.

Procedure at birth

For the **Network study** infants were presented with recordings of a fairy tale delivered in Hungarian in both IDS and ADS register by a native Hungarian-speaking mother, who at the time of recording was directing her words to her own 4.5-month-old infant or to the experimenter, respectively. Recordings were carried out in an anechoic room using a single sound channel and were later converted to dual-mono.

For the **Timing study** infants were presented with shorter (200 ms) or longer (300 ms) sounds grouped into two separate blocks each consisting of 150 short and 150 long sounds delivered in random order. One of the stimulus blocks presented tones of 500 Hz base frequency with 3 harmonics of 50, 33, and 25 percent amplitude, respectively, summed linearly together (‘tone’ condition). The other stimulus block comprised frozen white noise segments (‘noise’ condition). All sounds were presented at approximately 70 dB SPL loudness and the sounds were attenuated by 5 ms long raised-cosine onset

and offset ramps. The onset-to-onset interval was 800 ms. The tone condition was always delivered first, and there was a short rest of 30-60 seconds between conditions.
No stimulation was presented for the **Resting State Network** measurements.
Recordings were presented binaurally using the E-Prime stimulus presentation software (Psychology Software Tools, Inc., Pittsburgh, PA, USA) with ER-1 headphones (Etymotic Research Inc., Elk Grove Village, IL, USA) connected via sound tubes to self-adhesive ear cups (Sanibel Supply, Middelfart, Denmark) placed over the infants' ears. Stimuli were presented in two blocks. Within each block, the recording of each speech register was presented two times in a row, with an inter-stimulus interval of 10 s between recordings within the blocks (overall eight stimuli, 4 presentation of each recording, ca. 18 minutes duration). The order of the speech registers was counterbalanced between the blocks and randomized across infants. The experiment took place at the Department of Obstetrics-Gynecology and Perinatal Intensive Care Unit, Military Hospital, Budapest, with only the infant and the experimenter being present in the room.

Procedure at 4 and 9 months
For both the **Network study** and the **Timing study** the stimuli were identical to the newborn age recordings, however the very limited time available in these age groups necessitated to only present one stimulus/register (ca. 4.5 minutes overall duration). Stimuli were presented in a sound attenuated room using Matlab (MathWorks Inc., Natick, MA, USA) and Psychtoolbox (Kleiner, Brainard and Pelli, 2007). The sound signal from the computer was amplified by a Yamaha A-S301 amplifier (Hamamatsu, Japan) and presented through a pair of speakers at a comfortable intensity (Boston Acoustics A25, Woburn, MA, USA). The speakers were positioned ca. 1.75 meter in front of the participant, 70 cm apart. The experiment took place at the sound-attenuated and Faraday-shielded infant EEG laboratory of the Institute of Cognitive Neuroscience, Research Centre for Natural Sciences. The infant sat comfortably in his/her mother's lap while the experimenter employed toys to keep the infant facing towards the loudspeakers and his/her attention away from the electrode net. The mother was listening to music through closed can audiometric headphones to isolate her from the experimental stimulation.

EEG recording at birth
EEG was recorded during quiet sleep with Ag/AgCl electrodes attached to the scalp at the Fp1, Fp2, Fz, F3, F4, F7, F8, T3, T4, Cz, C3, C4, Pz, P3, P4 locations (location naming is in accordance with the international 10-20 system). The reference electrode was placed on the tip of the nose and the ground electrode on the forehead. EEG was digitized with 24 bit resolution at a sampling rate of 1 kHz by a direct-coupled amplifier (V-Amp, Brain Products GmbH, Germany). The signals were on-line low-pass filtered at 110 Hz. The impedance values were monitored throughout the recording session and were kept under 10 kΩ where possible, but no recordings were rejected based on impedance measurements. Although high impedance can introduce additional noise into the recordings, it may not be systematic in the entire group and its effects are mitigated by filtering out the very low frequencies, which is more likely to be contaminated by drifts. Infants' sleep state was determined based on behavioral criteria (Anders, Emde & Parmelee, 1971). Infants were predominantly asleep during the recordings, in quiet sleep 82% of the time, and in active sleep 8.2% of the time.

EEG recording at 4 and 9 months
EEG was recorded using a 60-channel HydroCel GSN net (64 channel v1.0 layout, channels 61-64 are connected to ground in the small pediatric caps used here) with an GES 400 DC amplifier passing the digitized signal to the NETSTATION v5.4.1.1 software (both Electrical Geodesics, Eugene, OR, USA). Signals were recorded online at a sampling rate of 1000 Hz with the Cz reference. Electrode impedance during recording was maintained below 50 kΩ. Electrodes corresponding to the Fp1, Fp2,

Fz, F3, F4, F7, F8, T3, T4, Cz, C3, C4, Pz, P3, P4 locations were selected for further analysis in order to keep the data compatible with the recordings in newborns.

Data analysis for ***Network study*** *and* ***Resting State Network***

EEG was analyzed with the EEGLAB (Delorme and Makeig, 2004) open-source Matlab (MathWorks Inc., Natick, MA, USA) toolbox. The signals were first bandpass filtered (using zero phase band-bass, Hamming windowed sinc FIR filter; filter order 33000) to the 0.5-45 Hz frequency range, then re-referenced to average of all electrodes. Continuous EEG data were segmented into 4096 ms long epochs with 50% overlap between successive epochs. Segments contaminated with physiological (eye movements, muscle artefacts) or external (i.e., environmental noise) artefacts were rejected by visual inspection. A minimum of 10 epochs was considered sufficient (Hillebrand et al., 2012; Tewarie et al., 2015) for further functional connectivity analysis. Epochs were then filtered (using zero phase band-bass, Hamming windowed FFT; filter order 1650) into five frequency bands (delta: (0.5)-4 Hz; theta: 4-8 Hz; alpha: 8-12 Hz; beta: 12-30 Hz; gamma: 30-(45) Hz;).
The strength of FC was calculated as pairwise phase synchronization strength between pairs of EEG electrodes, separately for the five frequency bands and for each epoch. The phase synchronization was measured by the PLI index (developed by Stam et al., 2007). PLI was calculated by using the BrainWave software version 0.9.151.5 (available at http://home.kpn.nl/stam7883/brainwave.html). PLI is an undirected measure of connectivity that calculates the consistency by which one signal is phase leading or lagging with respect to another signal.

Network analysis was performed on the graph theoretical representation of the functional connectivity matrix using the so-called minimum spanning tree (MST) approach (Boersma et al., 2013; Stam & van Straaten, 2012). The MST approach has been successfully employed for describing FC network properties (e.g. hierarchical structure, degree distribution etc.) of healthy newborn infants (Tóth et al., 2017). In the current study, the edge-weighted undirected graph consists of nodes that are the EEG electrodes (N=16) and edges (N=(16x16)/2-16) that are the connectivity values between each pair of nodes and the edge weights are the inverse PLI. MST is a subset of the edges (N=16) of a graph that connects all the nodes together without cycles (creating only one possible path between each pair of nodes) and with the minimum possible total edge weight (using the edges with the strongest FCs). MST networks were constructed by Kruskal's algorithm (Kruskal, 1956). MST was constructed separately for each epoch and frequency bands.
Graph measures described by Stam and colleagues (2014; see also Tewarie et al., 2015) were used. MST network characteristics values were normalized by dividing them by the number of EEG channels used. The global MST network characteristics were averaged across epochs, separately for each infant and frequency band. The level of integration of the communication within the network was assessed by the metric called “Diameter” (DIAM). DIAM is the largest distance between any two nodes within the MST, where distance refers to the minimum number of edges required to proceed from one node to another (the shortest path). The degree of segregation within a network is measured by the metric of “Leaf Fraction” (LF) and “Tree Hierarchy” (TH). Leaf Fraction” (LF) is the number of nodes with only 1 connected edge divided by the total number of nodes in the MST. “Tree Hierarchy” (TH) assesses how hierarchical a given network is compared to the so-called ‘star-like network organization’ (for a mathematical description, see Boersma et al., 2013 and Tewarie et al., 2015). TH ranges from 0 (indicating a line-like topology) to 1; for the star-like topology, TH approaches 0.5. The level of influence or centrality of the nodes within the network elements was assessed by the metric “Betweenness Centrality” (BC). BC is a measure of the node’s ‘hubness’ within the network. It is defined as the normalized fraction of all shortest paths connecting two nodes that pass through the particular node (Stam et al., 2014). BC is calculated separately for each node. BC may also be used to

identify regions of the brain that serve as hubs by plotting the distribution of BC over the scalp (see, e.g., Tóth et al., 2017).
The network measures averaged over epochs per stimulation (IDS, ADS, Resting State) frequency band (alpha, beta, gamma, delta, theta) and age (0, 4 and 9 months) were used as variables for the statistical analysis.

*Data analysis for **Timing study***
Signals were off-line filtered between 1–30 Hz and epochs of -100 to 800 ms with respect to the sound onset were extracted for each stimulus. The 100-ms pre-stimulus interval served as the baseline for amplitude measurements and illustrations. Due to low signal-to-noise ratio on other electrodes and comparability between age groups only data from electrodes F3, Fz, F4, C3, Cz, C4, P3, Pz, P4 were further analyzed. Epochs with a voltage change exceeding 120 µV on any channels measured in a moving window (window length = 100 ms, window step = 50 ms) in the -100-0 ms and 300-800 ms windows were rejected from analysis. Only data for infants that that had at least 100-100 trials in all conditions were used. Responses were averaged according to sound length (200 and 300 ms) and sound type (tone and noise) separately. Average amplitude measurement windows were selected based on maximum difference in grand average waveforms separately for the different ages and sound types: 0 month (200-300 ms noise/tone, 400-500 ms tone, 450-550 ms noise, 550-650 noise/tone) 4 month (200-300 ms noise/tone, 350-450 ms noise, 400-500 ms tone, 450-550 ms noise, 550-650 ms tone) and 9 month (300-400 ms noise, 350-450 ms tone, 400-500 ms noise, 450-550 ms tone, 500-600 ms noise). These averages were used as predictive variables in the statistical the analysis.

Statistical modeling methods
A machine learning pipeline utilizing Random Forests (Breiman, 2001) combined with MissForest imputation algorithm (Stekhoven and Bühlmann, 2012) and Boruta feature selection (Kursa and Rudnicki, 2010) was developed to build predictive regression models for each language outcome variable (CDI 18mo Receptive & Expressive, Bayley 18mo Receptive, Expressive, and Main scales) to reveal if 1) early maternal personality, maternal speech, and EEG features can predict language outcome at 18 months of age, 2) which of the predictor variables are statistically important, and 3) how those features influence the predicted outcome.

The pipeline was created to mitigate two major weaknesses of our data set in a consistent and statistically reliable manner: 1) The sample size was low (N=117), especially compared to the number of predictor variables (p=2011), and 2) the missingness ratio was large (50%) due to the high attrition rate. The Random Forest algorithm was chosen because it is applicable in such "low n, high p" problems (e.g., in genomic data analysis, see Chen and Ishwaran, 2012), and in most cases provides accurate predictions even with the default (untuned) parameters (Probst at al., 2019). The chosen imputation and feature selection algorithms (MissForest and Boruta, respectively) are wrappers around random forests, and perform very well compared to other imputation and feature selection algorithms (see e.g., Waelje at el., 2013, Degenhardt et al., 2019). To produce honest estimate of the generalization error and the stability of feature importance, leave-one-out cross-validation was applied so that in each cross-validation step, the given participant was held out from the sample, and the full pipeline (imputation, feature selection, final model on the restricted set of features) was run on the training sample and evaluated on the holdout participant (see Hastie et al., 2009). All analyses were carried out in R (R Core Team, 2022, version 4.2.1).

In particular, we repeated the following pipeline for each language outcome variable:
Prepare the data set:
Drop all columns which are not numeric or logical, and do not have at least 40 non-missing values;

Keep only those records for which at least one of the language outcome variables (9-18mo) and at least one of the predictor variables are non-missing;

Remove columns which have a single distinct value (e.g., all TRUE or FALSE);

For `CDI 18mo Expressive`, apply log-transformation on the target variable.

Apply leave-one-out cross-validation; that is, select all participants who have valid (non-missing) target values and for each of them, perform the following steps:

Split the data into a train set and a validation set, where the validation set is formed by the record of the given participant, and the train set contains all other records;

On the train set, do the following (always using the default settings of the given package, that is, without parameter tuning):

Impute the missing data by the missForest algorithm using the missRanger R package (Mayer, 2021). During the imputation, include all language outcome variables and all predictor variables, but after the imputation keep only the target variable and the predictor variables, and drop all records where the target variable was missing before the imputation (following von Hippel, 2007);

Select relevant features confirmed by the Boruta algorithm (as implemented in the Boruta R package, version 8.0.0). To remove the potential bias of highly unbalanced number of variables per variable categories (e.g., there are five Big 5 variables and hundreds of phonetic and EEG network variables), apply the Boruta algorithm in two rounds: In the first round, take each variable category separately, run the algorithm, and record the variables which were deemed important. In the second round, include all variables which were selected in the first round, and run the Boruta algorithm on them. Keep only those predictors which were found important in the second round. In addition to select the features for the final model, record the detailed feature importance statistics as calculated by the Boruta algorithm.

Run Random Forest algorithm (as implemented in the ranger R package: Write and Ziegler, 2017).

On the validation set, do the following:

Impute the missing values by the imputation model fitted to the train set in step 2.2.1.;

Predict the target variable by using the model fitted to the train set in step 2.2.3;

As a complementary baseline prediction, predict the target variable as the average of the target variable in the train set (that is, corresponding to an intercept-only linear model fitted to the train set).

Merge the results of each partition label (that is, the predicted scores for the validaton sets), and calculate performance measures (cross-validation R2 [cvR2]) and SHAP values (Lundberg and Lee, 2017) for feature interpretation (as implemented in the fastshap R package, Greenwell, 2021) on the validation set. cvR2 was calculated as

$$1-\frac{\sum_i(\hat{y}_i-y_i)^2}{\sum_i(\bar{y}_i-y_i)^2},$$

where y_i is the observed value of the target variable in the ith validation set, whereas $\hat{y}_i$ and $\bar{y}_i$ are the predicted values based on the fitted and the baseline models, respectively.

We also calculated another variant, the cross-validated relative absolute error [cvRAE] of the above measure using absolute instead of squared errors to decrease the effect of potential outliers.
Thus, cvR2 and cvRAE quantify the predictive accuracy of the final Random Forest model over a baseline model which simply uses the average of the observed target values in a given sample. Note that the upper bound of cvR2 and cvRAE as defined above equals 1, but they have no lower bounds as overfitting models can produce arbitrarily large squared or absolute errors in the validation sets, resulting in negative cvR2 and cvRAE.

If Step 3 indicates that the model performs better than the baseline model, calculate the SHAP Global Importance value (the average of absolute Shapley values across all participants) for each feature, and use also the detailed Boruta importance statistics (ratio of cross-validation folds when the feature was selected in the single-category and multi-category runs, average and median of mean importance values across all cross-validation runs) as auxiliary information to reveal the most important predictors.

If Step 3 indicates that the model performs better than the baseline model, use the individual SHAP values to reveal how a given feature contributed to the model prediction for each participant.

RESULTS

Initial analysis of the outcome measures

Descriptive data of the four main variables calculated from the CDI data at each time points is shown in Table 2. According to Shapiro-Wilk tests most variables do not show normal distribution. Test-retest reliability of the CDI measure was checked using Spearman correlation comparing receptive vocabulary scores recorded at each time points (Table 3). Strong and significant correlations between vocabulary scores measured at consecutive time points suggest the reliability of the measurements. Correlations gradually lose strength between further time points but are still moderate and significant between measurements taken at 9 and 18 months.

Table 2. CDI descriptives.

Measure	Receptive vocabulary					Gestures					Expressive vocabulary					Sentences				
Age (months)	**9**	**12**	**14**	**16**	**18**	**9**	**12**	**14**	**16**	**18**	**9**	**12**	**14**	**16**	**18**	**9**	**12**	**14**	**16**	**18**
N	96	95	90	90	86	96	95	90	90	86	96	95	90	90	86	96	95	90	90	86
Missing	23	24	29	29	33	23	24	29	29	33	23	24	29	29	33	23	24	29	29	33
Mean	49.1	141	208	268	310	17.1	31.3	41.6	50.2	57.8	2.21	6.77	16	33.6	65	7.8	17.6	21.3	24.4	25.1
Median	29.5	134	219	289	346	16	29	40	49.5	58.5	0	4	8	13.5	23	6	18	23	26	27
Standard deviation	61.5	97.6	110	106	107	7.1	9.9	12.7	12.5	12.3	3.85	9.74	23.6	53.8	96.4	5.94	6.35	5.78	4.88	4.42
Minimum	0	0	5	4	27	3	7	15	16	25	0	0	0	0	0	0	0	2	1	8
Maximum	391	447	447	455	455	35	62	80	79	81	20	45	116	274	383	25	28	28	28	28
Skewness	3.03	0.815	0.124	-0.528	-0.801	0.568	0.64	0.295	-0.105	-0.327	2.42	2.51	2.63	2.5	2.02	0.969	-0.472	-1.35	-2.24	-2.19
Std. error skewness	0.246	0.247	0.254	0.254	0.26	0.246	0.247	0.254	0.254	0.26	0.246	0.247	0.254	0.254	0.26	0.246	0.247	0.254	0.254	0.26
Kurtosis	12.2	0.392	-0.789	-0.534	-0.239	-0.0951	0.634	0.105	-0.0725	-0.218	6.34	6.59	6.94	6.16	3.14	0.374	-0.301	1.99	6.14	4.58
Std. error kurtosis	0.488	0.49	0.503	0.503	0.514	0.488	0.49	0.503	0.503	0.514	0.488	0.49	0.503	0.503	0.514	0.488	0.49	0.503	0.503	0.514
Shapiro-Wilk W	0.691	0.939	0.973	0.959	0.922	0.965	0.969	0.985	0.992	0.983	0.642	0.673	0.645	0.62	0.651	0.915	0.968	0.883	0.735	0.687
Shapiro-Wilk p	< .001	< .001	0.061	0.007	< .001	0.012	0.022	0.375	0.853	0.318	< .001	< .001	< .001	< .001	< .001	< .001	0.022	< .001	< .001	< .001
10th percentile	5	26.2	61.9	99.8	153	9	20	25	35	43.5	0	0	0	1	3	2	9	13.9	17.9	19.5
25th percentile	11	58.5	125	200	233	13	25	33	42	48.5	0	1	3	5	12	3	13	19	23	25
50th percentile	29.5	134	219	289	346	16	29	40	49.5	58.5	0	4	8	13.5	23	6	18	23	26	27
75th percentile	66.3	201	270	348	389	22	35.5	50.8	58	66.8	3	8	16.8	26.8	48.5	11	22.5	25	28	28
90th percentile	111	264	360	389	417	27.5	45.4	58.2	66	73	6.5	17	33.7	120	228	16	25	27.1	28	28

Table 3. Results of Spearman correlations comparing receptive vocabulary scores measured at consecutive time points.

Correlation matrix		receptive_9m		receptive_12m		receptive_14m		receptive_16m	
receptive_12m	Spearman's rho	0.719	***	—					
	p-value	<.001		—					
receptive_14m	Spearman's rho	0.618	***	0.895	***	—			
	p-value	<.001		<.001		—			
receptive_16m	Spearman's rho	0.568	***	0.803	***	0.871	***	—	
	p-value	<.001		<.001		<.001		—	
receptive_18m	Spearman's rho	0.535	***	0.704	***	0.748	***	0.817	***
	p-value	<.001		<.001		<.001		<.001	

The age-related growth of receptive and expressive vocabularies, communicative gesture use and sentence comprehension was analysed using non-parametric Friedman's test showing a statistically significant increase across time points in Receptive vocabulary, $\chi2(4) = 284$, $p < 0.001$, Expressive vocabulary, $\chi2(4) = 215$, $p < 0.001$, Gestures, $\chi2(4) = 305$, $p < 0.001$, and Sentences, $\chi2(4) = 269$, $p < 0.001$. Pairwise comparisons using Durbin-Conover tests revealed significant growth in each measure between consecutive time points (Table 4).

Table 4. Results of pairwise comparisons using Durbin-Conover tests for each variable measured at consecutive time points.

Receptive vocabulary				Expressive vocabulary			
Time points		Statistic	p	Time points		Statistic	p
9m	12m	12.74	<.001	9m	12m	5.41	<.001
12m	14m	11.34	<.001	12m	14m	4.8	<.001
14m	16m	10.08	<.001	14m	16m	5.46	<.001
16m	18m	7.91	<.001	16m	18m	7.17	<.001

Gestures				Sentences			
Time points		Statistic	p	Time points		Statistic	p
9m	12m	16.8	<.001	9m	12m	11.83	<.001
12m	14m	16.1	<.001	12m	14m	9.75	<.001
14m	16m	16.4	<.001	14m	16m	9.56	<.001
16m	18m	13.7	<.001	16m	18m	2.39	0.017

The growth of language abilities measured by the CDI variables is illustrated in Figures 1-3.

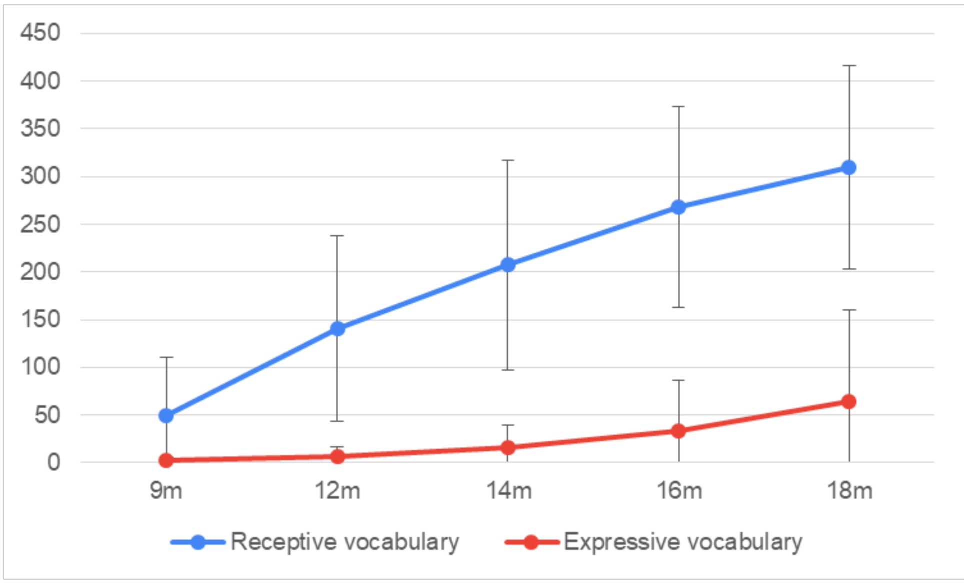

Figure 1. Average CDI-I Receptive and Expressive vocabulary scores by age.

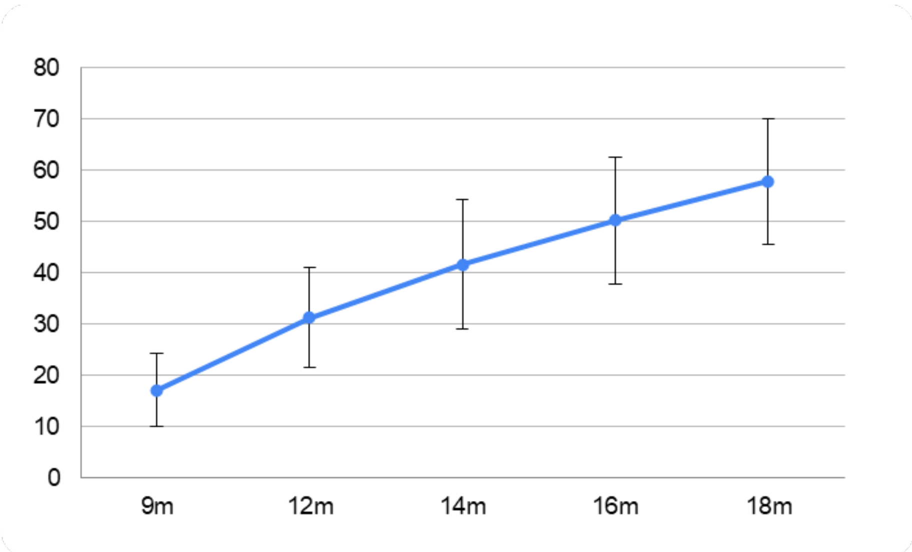

Figure 2. Average CDI-I Gestures scores by age.

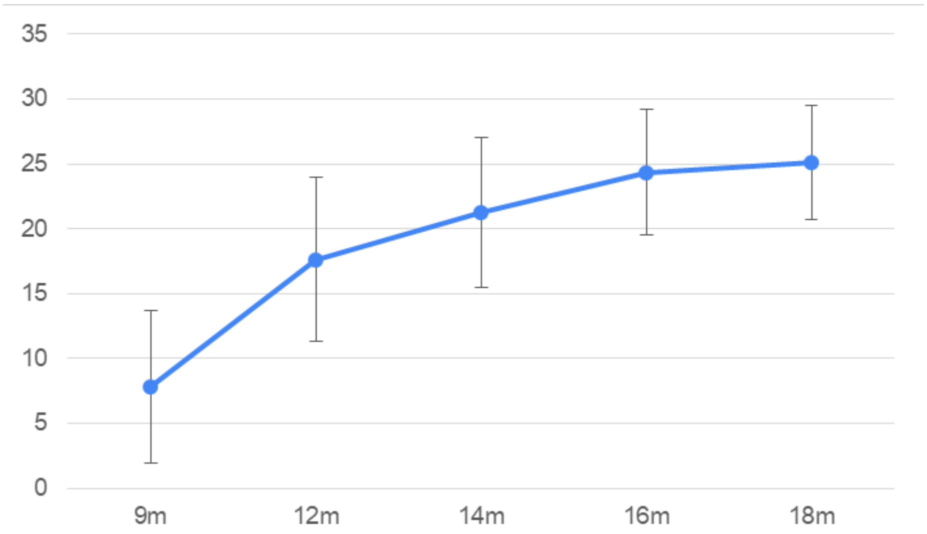

Figure 3. Average CDI-I Sentence comprehension scores by age.

Correspondences between expressive and receptive vocabulary development

The correspondence between CDI-I expressive and receptive vocabulary scores was analysed using Spearman rank correlations. Correlation was weak at 9 months, r=0.306, p< .01, while being moderate at 18 months, r=0.454, p< .001. Although increasing correlations show that the correspondence between receptive and expressive vocabularies strengthen with age in this age range, there is a great variation in both scores. The scatterplot in Figure 4 illustrates that (i) considerable expressive vocabulary scores are only observable in infants with relatively high receptive vocabulary scores, but (ii) many infants with high receptive vocabulary scores are still not using more than a few words expressively if any.

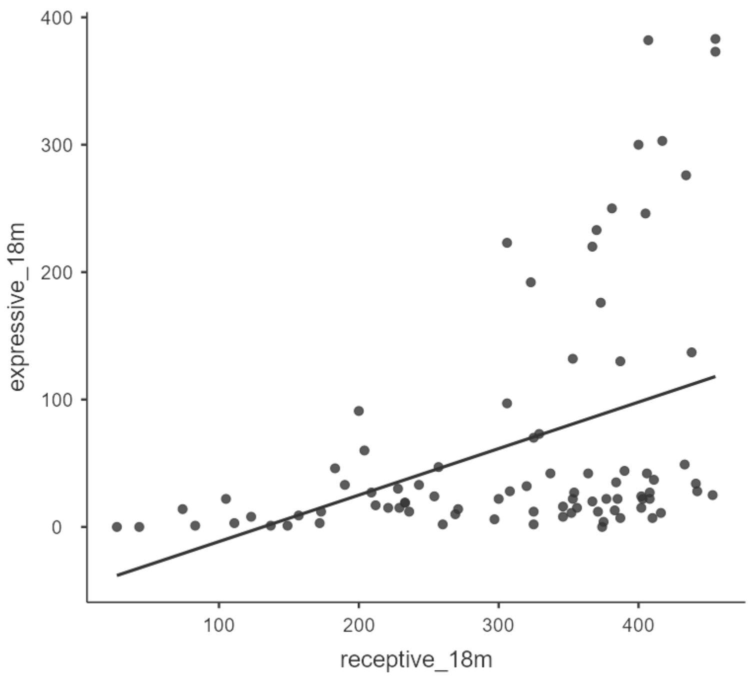

Figure 4. Correspondences between CDI-I Receptive and Expressive vocabulary scores in individual participants at 18 months.

Classification of participants with different language outcomes

In order to identify subgroups of infants based on outcomes of language development we used K-means clustering that separated 3 clusters of children based on all four CDI-based variables recorded at 18 months (Receptive and Expressive vocabulary, Gestures, Sentences). Gap statistic was highest in case of three clusters within the range based on four variables (1-3 clusters; Table 5).

Table 5. Centroids of clusters based on CDI variables at 18 months (standardized z-scores).

Cluster No	Cluster count (N)	receptive_18m	expressive_18m	sentences_18m	gestures_18m
1	13	0.767	2.165	0.543	0.773
2	50	0.412	-0.330	0.421	0.250
3	23	-1.330	-0.507	-1.222	-0.980

The plot of variable means (converted to standardized z-scores) reveals that k-clustering separated three groups of infants who are characterized by different performances in certain domains of language. The highest performing children (Cluster 1, N=13) achieved similar mean scores in most outcome variables (receptive vocabulary, sentence comprehension and gestures) to children in the largest cluster (Cluster 2, N=50) showing average performance in all variables. The only significant mean difference (>1 SD) between Cluster 1 and 2 favoring the former group is Expressive vocabulary. The lowest performing children (Cluster 3, N=23) achieved similar mean scores in expressive vocabulary to children in the largest cluster (Cluster 2) but performed significantly below them in all remaining outcome variables (receptive vocabulary, sentence comprehension and gestures). Thus, children in Cluster 1 stand out from the sample by showing an extraordinary expressive vocabulary while infants in Cluster 3 lag behind in receptive language abilities and gesture use. (Figure 5).

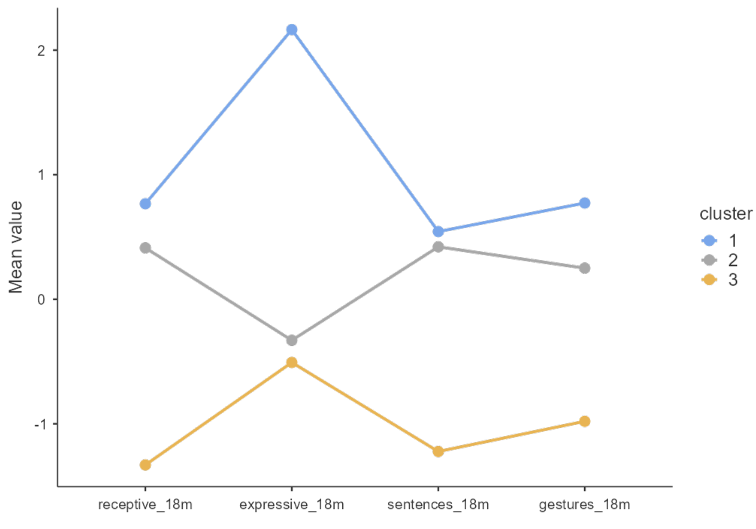

Figure 5. Plot of variable means across clusters (standardized scores).

However, this grouping of children based on early language achievements cannot be considered strongly predictive of later language outcomes. According to the literature, developmental trajectories in children with language delay are highly variable and there is a significant inconsistency between measurements at different time points. Statistically based approaches categorize children with expressive vocabulary below 10th or 15th percentile and producing no word combinations as late talkers (Thal and Bates 1988, Fenson et al. 1993). At later time points, Dale et al (2003) identified 10.7% of 3-year-old and 11.5% of 4-year-old children as showing language delay (LD) based on parent report (CDI) using the 15th centile criterion in at least 2 out of 3 subparts measuring mainly expressive vocabulary and communication. Dale et al. (2003) compared measurements taken at 24, 36 and 48 months and reported that less than half of the children with early language delay at 2 years show persistent language difficulties one or two years later (44.1% at 3 years and 40.2% at 4 years) with most of these children catching up to the normal range spontaneously in one or two years. They also found that most of the children showing low language capabilities at 3 or 4 years have not been classified as language delayed at 2 years, referring to the low predictability of language outcomes based on early assessments. This pattern of findings is supported by other studies as well (Feldman et al. 2005, Westerlund et al. 2006 in Swedish, Lyytinen et al. 2001 in Finnish).

Henrichs et al. (2011) distinguishes between 3 different subtypes/trajectories of persistent, transient and late-onset language delays in their large-sample longitudinal study with measurements taken at 18 and 30 months in Dutch language. The corresponding proportions of children falling into these three categories are 6.2% with transient LD (low expressive vocabulary at 18 but normal scores at 30 months), 6.0% with late onset expressive vocabulary delay (normal expressive vocabulary at 18 but below cut-off scores at 30 months), and 2.6% with persistent expressive vocabulary delay (at both time points) in the whole sample. A prediction model including demographic and familial factors along with early receptive and expressive vocabulary scores could only explain 17.7% of the variance in 30-month expressive vocabulary with sensitivity and positive predictive values for the CDI–N being very low.

Our grouping, however, is different from those cited above as we included receptive vocabulary and sentence comprehension scores as well along with gestural communication to the variable set on which the grouping of children was based. Furthermore, the assignment of groups was not determined by a particular percentile, but the cluster analysis separated children based on the distribution of the data. Further follow-up of the children will explore the consistency of this clustering in light of the later language achievements of the same participants.

Description and scoring of the Bayley-III Language Scale

The Bayley-III Language Scale consists of tasks measuring Receptive and Expressive Communication, which form separate subtests. Receptive communication (RC): 49 test items measuring preverbal behaviour; vocabulary development (e.g., identification of objects or pictures presented); morphological vocabulary development (e.g., knowledge of nouns and prepositions); and understanding of morphological markers, plurals, verb tenses and possessive relations. The scale also includes tasks that test social reference and verbal comprehension. Expressive communication (EC): 48 test items measuring preverbal communication (e.g., babbling), gesticulation, shared reference, and speech shift; vocabulary development (e.g., naming objects, pictures or attributes [e.g. colour, size]); and morphosyntactic development, use of two-word phrases, plural and verb tenses.

Raw scores of the RC and EC tasks are converted to RC and EC standard scores based on age-dependent population norms. The sum of the RC and EC standard scores are then transformed yielding the Language Composite Score. For different analyses we used the RC and EC standard scores as well as the Language Composite Score at 9 and 18 months of age. Descriptive data of the Expressive and Receptive communication standard scores and the Language Composite Score at 9 and 18 months is shown in Table 6. It is apparent that Language Composite Scores of our sample show close similarity to the population norm at 18 months (mean=100, SD=17.6). However, this is not the case at 9 months where the present sample's mean is below the population norm with 1 SD (mean=85.5, SD=12.6).

Table 6. Bayley-III Language Scale descriptives.

	9 months			18 months		
	Bayley-III Receptive	Bayley-III Expressive	Bayley-III Language Composite	Bayley-III Receptive	Bayley-III Expressive	Bayley-III Language Composite
N	80	80	80	66	66	66
Missing	37	37	37	51	51	51
Mean	7.84	7.15	85.5	11.5	8.56	100
Median	8.00	7.00	87.5	13.0	8.00	103
Standard deviation	2.94	2.53	12.6	4.40	2.74	17.6
Minimum	2	1	56	2	2	53
Maximum	17	12	121	18	16	138
Skewness	0.173	-0.139	0.0388	-0.877	0.525	-0.554
Std. error skewness	0.269	0.269	0.269	0.295	0.295	0.295
Kurtosis	0.176	-0.545	-0.154	-0.216	0.669	0.275
Std. error kurtosis	0.532	0.532	0.532	0.582	0.582	0.582
Shapiro-Wilk W	0.969	0.974	0.984	0.892	0.959	0.961
Shapiro-Wilk p	0.049	0.109	0.446	<.001	0.027	0.036
10th percentile	4.00	4.00	71.0	4.50	5.50	75.5
25th percentile	6.00	5.00	77.0	8.50	7.00	91.0
50th percentile	8.00	7.00	87.5	13.0	8.00	103
75th percentile	10.0	9.00	94.0	15.0	10.0	112
90th percentile	11.0	10.0	100	15.5	11.5	118

According to Shapiro-Wilk tests most variables do not show normal distribution. Therefore, differences of receptive and expressive communication scores and the language composite scores between 9 and 18 months was checked using non-parametric Wilcoxon's test showing a statistically significant increase from 9 to 18 months in Receptive communication standard score, Expressive communication standard score and Language Composite Score (Table 7). Note that only 54 infants were successfully tested with the Bayley-III at both time points.

Table 7. Results of the Paired Samples Wilcoxon Tests comparing Bayley-III scores measured at 9 and 18 months.

			Statistic	p	Mean difference	SE difference	Effect Size
Bayley-9m-rec-standard	Bayley-18m-rec-standard	Wilcoxon W	204	<.001	-3.50	0.635	-0.653
Bayley-9m-exp-standard	Bayley-18m-exp-standard	Wilcoxon W	320	0.025	-1.50	0.424	-0.383
Bayley-9m-lang-composite	Bayley-18m-lang-composite	Wilcoxon W	198	<.001	-12.81	2.415	-0.701

Correspondences between different measures of language development

The comparison between different data sources on language development – parent report CDI and Bayley-III testing – was analyzed using Spearman rank correlations. Analyses showed lack of correlations between the Bayley-III Language measures and the CDI-scores at 9 months with only a weak and significant correlation between CDI Gestures score and Bayley-III Expressive Communication score. At 18 months, however, significant and weak to moderate correlations are observed between several variables coming from the two the different data sources (Table 8-9).

Table 8. Spearman correlations between Bayley-III Language measures and the CDI-scores at 9 months.

		CDI-receptive		CDI-gestures		CDI-expressive		CDI-sentences	
Bayley-Rec-standard	Spearman's rho	0.105		0.053		0.034		0.105	
	p-value	0.352		0.638		0.768		0.355	
Bayley-Exp-standard	Spearman's rho	0.060		0.279	*	0.175		0.139	
	p-value	0.599		0.012		0.120		0.219	
Bayley-Lang-Composite	Spearman's rho	0.106		0.199		0.122		0.150	
	p-value	0.349		0.077		0.281		0.183	

Table 9. Spearman correlations between Bayley-III Language measures and the CDI-scores at 18 months.

		CDI-receptive		CDI-gestures		CDI-expressive		CDI-sentences	
Bayley-Rec-standard	Spearman's rho	0.418	***	0.428	***	0.338	**	0.307	*
	p-value	<.001		<.001		0.006		0.013	
Bayley-Exp-standard	Spearman's rho	0.119		0.283	*	0.706	***	0.162	
	p-value	0.347		0.022		<.001		0.198	
Bayley-Lang-Composite	Spearman's rho	0.357	**	0.428	***	0.619	***	0.288	*
	p-value	0.004		<.001		<.001		0.020	

The strongest significant correlations between CDI and Bayley scores at 18 months are those between CDI Expressive Vocabulary score and Bayley-III Expressive Standard Scores ($r=0.706$, $p < .001$), and between CDI Expressive Vocabulary and Bayley-III Language Composite Score ($r=0.619$, $p < .001$). This might reflect the fact that parent's judgments on their children's language abilities are more reliable and valid in the domain of expressive communication compared to receptive abilities.

Concurrent validity of the language outcome measures applied in our research project was also analyzed by comparing Bayley-III Language scores measured at 18 months in the three groups identified using K means clustering based on CDI variables (see above), descriptive data is shown in Table 10. There were 65 children out of 86 who were accessible for the Bayley-III testing during the research project.

Table 10. Mean Bayley-III Language scores by group based on CDI-scores at 18 months.

	Group (CDI-clustering)	N	Mean	SD
Bayley_18m_RC	1	12	13.50	3.18
	2	37	12.65	3.65
	3	16	8.06	4.54
Bayley_18m_EC	1	12	12.67	1.97
	2	37	7.76	1.82
	3	16	7.38	2.25
Bayley_18m_LC	1	12	118.17	13.33
	2	37	101.38	12.56
	3	16	86.81	18.47

Kruskal-Wallis ANOVA showed significant effect of Group for Bayley-III Expressive communication scores, $\chi^2(2)=28.2$, $p<.001$, $\varepsilon^2=0.440$, Receptive communication scores $\chi^2(2)=15.2$, $p<.001$, $\varepsilon^2=0.238$, and the Language Composite Score, $\chi^2(2)=20.5$, $p<.001$, $\varepsilon^2=0.320$. Dwass-Steel-Critchlow-Fligner pairwise comparisons revealed that

(i) Group 3 perform significantly below Group 2 Bayley-III Receptive communication score, $W=5.14$, $p<.001$, while Group 1 and 2 do not differ, $W=1.12$, $p=.710$,

(ii) Group 1 significantly outperforms Group 2 in Bayley-III Expressive communication score, $W=7.195$, $p<0.001$, while Group 2 and 3 do not differ, $W=0.584$, $p=.910$,

(iii) Group 1 significantly outperforms Group 2, $W=4.74$, $p=.002$, and Group 3 perform significantly below Group 2, $W=3.84$, $p=0.018$, in Bayley-III Language Composite Score.

Bayley-III language scores show the pattern to be expected based on CDI scores: Groups 1 and 2 differ only in expressive vocabulary on the CDI and Expressive communication score and Language Composite Score on the Bayley-III, while Groups 2 and 3 differ on receptive vocabulary, sentence comprehension and gestures on the CDI and Receptive communication score and Language Composite Score on the Bayley-III. Thus, the two different data sources show highly concurrent results in terms of linguistic levels and language use, supporting concurrent validity of the language outcome measures used in our research.

Statistical modeling

Out of the five target outcomes, the final models improved the generalized prediction accuracy only for the CDI receptive vocabulary scores measured at the age of 18 months upon the baseline (intercept-only) model (see Table 11). For all other outcomes, our pipeline resulted in overfitted models.

Table 11. Generalized (cross-validated) accuracy of the final models.

Target Variable	cvR^2	cvRAE	Pearson r	Spearman ρ
CDI Receptive (18m)	0.178	0.123	0.400	0.370
CDI Expressive (18m)	-0.100	-0.111	0.010	-0.050
Bayley Receptive (18m)	-0.300	-0.074	-0.180	-0.200
Bayley Expressive (18m)	-0.030	0.012	0.070	0.150
Bayley Main (18m)	-0.240	-0.058	-0.180	-0.110

In the followings, we report only on the results of the models predicting CDI receptive vocabulary.

Regarding the importance of individual predictor variables, in an ideal scenario, the Boruta algorithm would have selected the very same set of predictors in every cross-validation run. As expected, this was not the case, so we got partially overlapping, and not identical predictor sets. Nevertheless, some predictors (age of mother at child's birth, BIG5 Extraversion, F2 frequencies of the vowels in infant-directed speech (IDS) relative to those in adult-directed speech (ADS)) were selected in >75% of the cross-validation runs, and even a particular EEG measure (Theta…) was confirmed by Boruta in two thirds of the CV runs (see Table 12).

Table 12. Variable importance (top 20 according to SHAP Global Importance).
Boruta Confirmed (in Single-Category): Ratio of cross-validation folds when the variable was selected in the single-category run;
Boruta Confirmed (in Multi-Category): Ratio of cross-validation folds when the variable was selected in the multi-category run;
Boruta Mean Importance: The average of Boruta importance values across all cross-validation folds;
SHAP Global Importance: The average of absolute Shapley values across all participants.

			Boruta			
Variable	Category	Age (month)	Confirmed (in Single-Category)	Confirmed (in Multi-Category)	Mean importance	SHAP Global Importance
phon__id_ad_frm_f2_m_6m	phon	6	0.78	0.77	4.51	9.43
ses__0m__m_eletkor	ses	0	1.00	0.99	5.96	8.44
phon__id_ad_gnl_en_min_6m	phon	6	0.81	0.81	4.67	6.76
big5__6m__extraversion	big5	6	1.00	0.90	3.97	5.81
eegfull__theta_4ho_ad_bc_ch_13	eegfull	4	0.69	0.63	3.73	4.83
phon__id_ad_gnl_en_iqr_6m	phon	6	0.52	0.52	2.42	4.06
phon__id_glob_rng_m_6m	phon	6	0.52	0.52	3.01	3.78
phon__id_ad_gnl_en_m_6m	phon	6	0.48	0.47	2.27	1.89
eeg_rs__gamma_4h_bc_ch_7	eeg_rs	4	0.31	0.23	1.16	1.83
phon__id_gnl_en_qm_0m	phon	0	0.21	0.20	1.20	1.80

phon__id_ad_gnl_en_c0_6m	phon	6	0.43	0.43	2.06	1.59
phon__id_ad_gnl_en_rms_6m	phon	6	0.20	0.17	0.97	1.20
phon__id_ad_gnl_en_med_6m	phon	6	0.40	0.38	1.99	1.13
eegfull__gamma_4ho_id_th	eegfull	4	0.26	0.19	1.06	1.09
eegfull__gamma_4ho_id_bc_ch_13	eegfull	4	0.09	0.05	0.30	0.88
phon__id_ad_gnl_f0_m_0m	phon	0	0.10	0.10	0.49	0.86
eeg_rs__gamma_9h_degree_ch_14	eeg_rs	9	0.16	0.16	0.71	0.81
eeg_rs__beta_4h_bc_ch_7	eeg_rs	4	0.13	0.09	0.43	0.67
timing_erp__0m_550_650_bin3_200_noise_pz	timing_erp	0	0.13	0.07	0.41	0.66
phon__id_glob_ml_bl_cross_t_0m	phon	0	0.03	0.03	0.19	0.64

It was also apparent that the SHAP Global Importance values were highly correlated to the Boruta importance statistics (Pearson r = 0.94 and 0.945 for the average and median Boruta importance vs SHAP, respectively).

See Figure 6 to explore how the top 20 most important features influenced the final models' predictions. E.g., participants whose mothers' speech was characterized by larger second formant frequencies (F2) of the vowels in infant-directed speech (IDS) relative to those in adult-directed speech (ADS) were predicted to have higher CDI Receptive Vocabulary score at the age of 18 months, whereas relatively higher age of the mother contributed to lower predicted scores. Figure 7 provides another, even more detailed view of the same results.

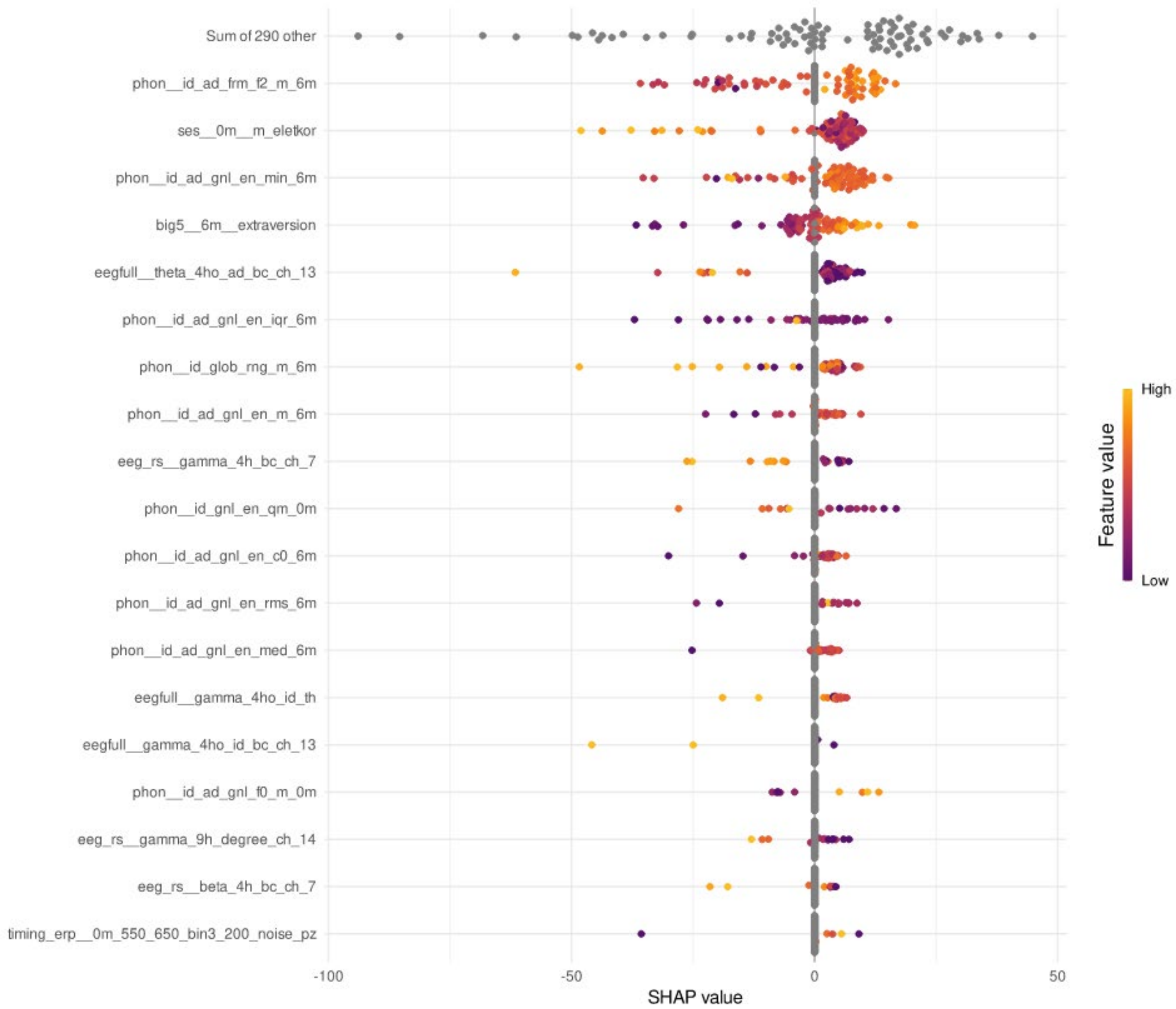

Figure 6. Top 20 most important features influencing the final models' prediction.

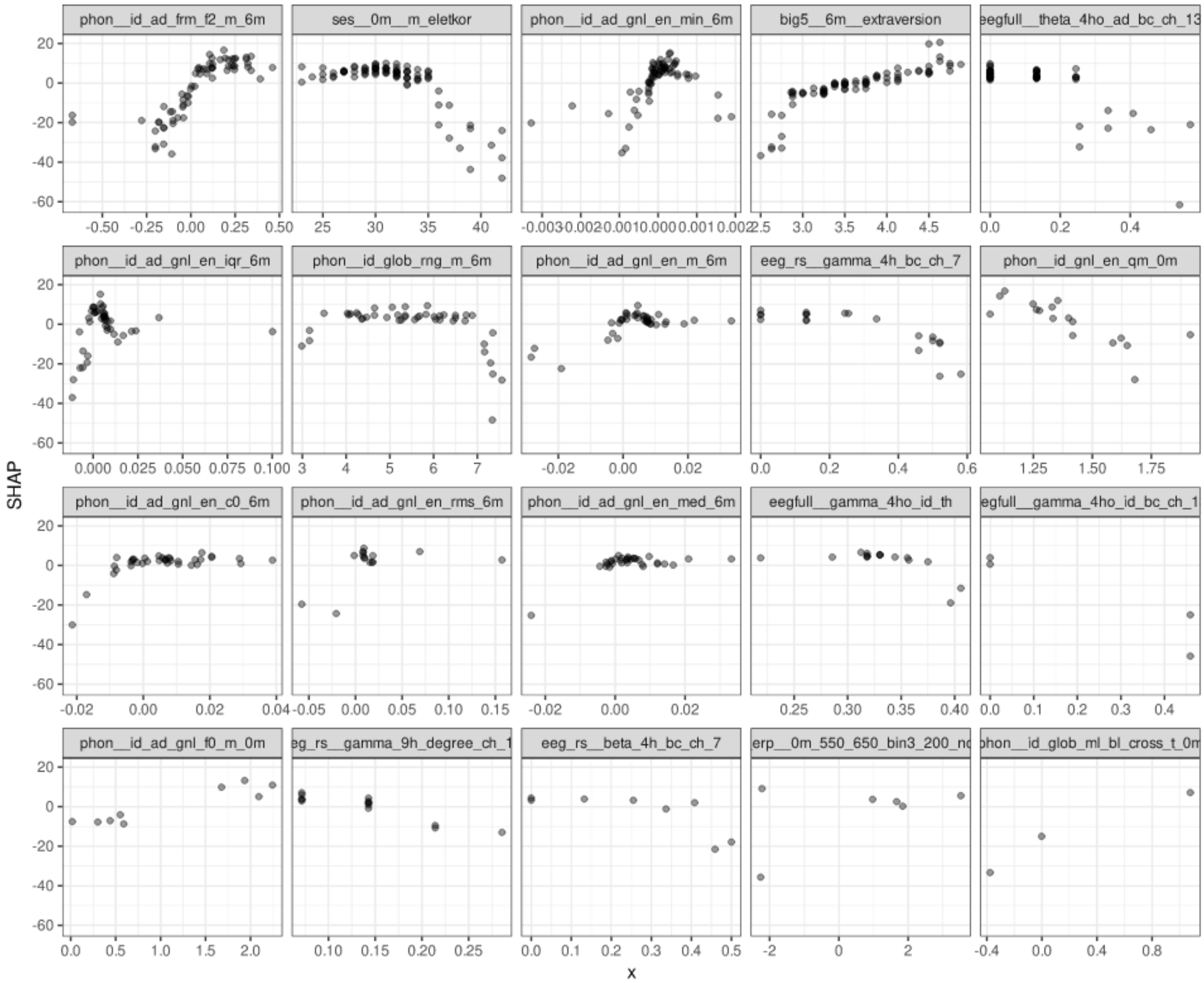

Figure 7: Original values of the predictors versus estimated SHAP Values for each participant and each of the top 20 most important predictors.

Discussion

Application of machine learning methods to predict early language outcomes

Our findings demonstrated that modern machine learning methods can be successfully applied to answer research questions and reveal new insights with relation to early language development, a field dominated by classical statistical methods. In particular, by analyzing the joint set of predictors in a transparent and data-driven fashion and scoring the models on previously unseen data (via cross-validation), instead of cherry-picking some variables and running correlation-based analyses on them is a valid and reliable complementary approach to standard model-based, hypothesis-driven experiments. We showed that from a very broad set of potential predictors a few variables consistently emerged as important precursors of the 18-month receptive vocabulary reported by the parent (CDI). The overall accuracy of the final model was moderate, though, which was expected given that our measures are only distant proxies of early language development, and individual trajectories of early language development themselves are known to be highly variable and hard to measure. The latter problems, that is, challenges in the reliable measurement of early expressive vocabulary probably contributed to the negative results that we found when predicting expressive language outcomes, in addition to the low sample size and high missingness ratio which are known to increase the probability of overfitting in most machine learning techniques.

Effects of maternal age and personality

The final model predicting 18-month receptive vocabulary as measured by the CDI yielded maternal age as one of the salient predictors. Our finding is in contrast with most reports showing positive relation between mothers' age and child early language skills. Older maternal age, which is usually associated with more years spent in school, hence implicating higher educational level, has been linked to better infant language skills (e.g., Bornstein et al. 2020). In a study of a large Australian community sample however infants' lower communication abilities measured at 24 months were predicted by older maternal age, but not educational level (Reilly et al., 2007). While Reilly et al. would not elaborate on this finding, we explored our data based on the SHAP effect visualization, and two maternal age groups (< 35ys vs. ≥ 35ys) were distinguished for further analyses. None of the demographic, birth and infant variables showed significant difference between the groups. Regarding maternal personality, Agreeableness scale scores were significantly lower (MW $Z=-2.224$, $U=655.5$, $p=.026$, $d_{Cohen}=.437$) in the older group ($M_{22}=3.70\pm.54$) than in younger mothers, under age 35 ($M_{86}=3.96\pm.53$). Thus, we might speculate that the specificity of the older group of mothers lies in their personality characteristics such as lower level of cooperativeness, trust, and warmth.

Another personality factor emerged as a salient predictor in our final model: infants of mothers scoring higher on the Extraversion scale showed better language understanding, i.e., more socially vital, actively engaging mothers had infants with richer 18-months receptive vocabulary. The two findings converge and suggest that maternal personality – through its influence on the quality of parenting – can play a role in early language acquisition. A meta-analytic review (Prinzie et al., 2009) investigating the association between the Big Five personality factors and three dimensions of parenting (warmth, behavioral control, autonomy support) revealed modest, but meaningful effects with child age as a moderating factor. For younger children and parents, stronger relation between Agreeableness and parental warmth was shown which is the age when warmth appears a dominant aspect of adaptive parenting. Madigan et al. (2009) conducted two separate meta-analyses examining the relations between early language development and parental warmth, as well as language and parental sensitive responsiveness in typically developing populations. They concluded that the association was stronger

for parental sensitivity than warmth, although both parental behaviors significantly predicted child language outcome.
Regarding our findings on the relation between maternal personality factors and infant receptive vocabulary, we assume that Extraversion expressed as more social engagement, more positive affect, higher activity level and more speech during mother-infant interactions could lead to a higher proportion of beneficial contexts for the infant to learn about events and objects. High Agreeableness can be reflected in warm, supportive, and sensitive interpersonal behavior, while its low level can lead to an insensitive, detached, or intrusive/hostile interactive style, hindering infants' learning in social situations. Underpinning this plausible explanation, a study with 20-month-old firstborn infants (Bornstein at al., 2011) found that parental Extraversion was associated with self-reports of parental competence and the amount of social interactions and observed symbolic play with the child. They also showed that maternal Agreeableness was associated with mothers' speech (MLU and diversity of words roots used) measured during interactions.

Effects maternal speech variables
Three major groups of acoustic parameters of the mothers' speech were found to be related to the later vocabulary of the babies: the formant frequencies of the vowels, speech intensity, and features connected to fundamental frequency. The most important parameters to influence the model's output were the second formant frequencies (F2) of the vowels in infant-directed speech (IDS) relative to those in adult-directed speech (ADS). The higher these F2-values were in IDS with respect to ADS, the more typical it became for the babies to develop larger vocabularies by the age of 18 months. These higher values of F2 imply that the mothers produced vowels more to the front in IDS, stimulating the babies' speech development. These findings are in agreement with results from previous research, which also revealed a connection between the formant space of the vowels and the later vocabulary (Hartman et al., 2017; Kalashnikova & Burnham, 2018). One possible explanation for this phenomenon is that in IDS the vowels are typically formed farther away from each other making them more easily detectable for the baby, thus the production of the vowels may facilitate the recognition of words. Besides, F2 values are connected to more lip spreading, among others due to smiling (Robson & Beck 1999, Barthel & Quené 2015), and this kind of emotion expression enhances the baby's speech development.

Effects of EEG variables
The variables related to expressive vocabulary of infants cover a mixture of EEG techniques, age groups, network measures, stimulus conditions, frequency bands and topological scalp positions making it extremely difficult to single out variables for interpretation. However the pattern apparent in the variable is consistent with our current knowledge of brain development and speech development in particular. Of the top seven variables six has topographical information connecting the variables to frontal and parietal channels. Studies show that already at birth infants recruit frontal areas during processing speech sound (Dehaene-Lambertz, Dehaene & Hertz-Pannier, 2002; Dehaene-Lambertz et al., 2002; Gervain et al., 2008). Moreover, temporal and frontal regions do not develop independently but show high correlations in their developmental course (Leroy et al., 2011). These findings suggest that the frontal cortices and their connections to auditory sensory regions (temporo-parietal areas) play a major role in language acquisition. In infants, frontal and temporo-parietal regions are already well connected and these connections are hypothesized to underlie the early processing of speech (Dubois et al., 2016).
Furthermore, the frequency bands found among the variables, namely Beta, Gamma and Theta are all related to language processing. Brain oscillations in the theta range synchronize to the dynamics of the speech envelope associated with syllabic (~5 Hz) and phrasal (~2 Hz) rates presentation respectively, while neural activity in the beta and gamma range follow the fine-grained temporal dynamics related to phonetic features (~20 Hz; Giraud and Poeppel, 2012; Leong and Goswami, 2014; 2015). Therefore

the networks operating in these frequency ranges are of general interest to speech perception and also specifically to processing IDS and ADS.
Of all ages the prevalence of 4 month can be interpreted as a shift occurring in speech processing networks around 4 months related to important developmental changes taking place (e.g., stronger preference for social sounds, McDonald et al., 2019 and the ability to produce vowel-like sounds and babble (Kuhl, 2004)). The variables here are related to the measure Betweenness Centrality (BC) which is a measure of "hubness", centrality within a network in this case emerging in frontal ad parietal regions. Apart from language related changes reflected in both AD and ID stimulus conditions appearing, the age of 4 months is also important for general brain development towards structural and functional integration (Zhao, Xu & He, 2019) as BC in resting state without stimulation is also seems to be important.

Limitations, conclusions, and future directions
The main limitation of the current investigation stems from the fragmentary nature of the data base. Largely due to the Covid 19 pandemic, it was not possible to collect a homogeneous data set, following the same infants from 0 to 18 months of age. Therefore, when measuring could be resumed, we included groups of infants entering the study at later ages, which allowed collecting only a subset of the predicting variables. Together with the normal rate of attrition and missed or unsuccessful measurements, this caused a very large amount of missing data in the final data set. While imputing missing data was very carefully executed, the lost information reduced the reliability of the statistical result. Using several steps of cross-validation, we could reasonably ascertain the findings. However, it is quite possible that the data quality was insufficient for discovering other predictive relationships. Further, the lower than planned data quality prevents us from studying interactions between the different factors.

The measurements and variables found to be most predictive of the outcome variables should serve as indicators for future studies, allowing them to focus more closely on certain aspects of the prerequisites of language acquisition as well as for early signs of potential problems for language development. One important direction is to study how the different factors interact with each other and degree to which early intervention affecting one can compensate for deficits in other factors.

Summarizing our results, we found that all main groups of studied variables (socioeconomic status, maternal personality and speech quality, and neural development) impacted on the infants' language acquisition, demonstrating that the complexity and sensitivity of this process. Each of these findings is (separately) compatible with results of other studies, supporting their validity. Thus, beyond the fact that language acquisition can indeed be predicted to a certain degree from early measurements, the current study adds the important message to the growing literature of language acquisition that these factors should be studied together.

References

Anders, T., Emde, R., & Parmelee, A.A. (1971). Manual of standardized terminology, techniques and criteria for scoring of states of sleep and wakefulness in newborn infants. Los Angeles, CA: UCLA Brain Information Service, NINDS Neurological Information Network.

Barthel, H., & Quené, H. (2015). Acoustic-phonetic properties of smiling revised: Measurements on a natural video corpus. In *Proceedings of the 18th international congress of phonetic sciences*.

Boersma, M., Smit, D. J., Boomsma, D. I., De Geus, E. J., Delemarre-van de Waal, H. A., & Stam, C. J. (2013). Growing trees in child brains: graph theoretical analysis of electroencephalography-derived minimum spanning tree in 5-and 7-year-old children reflects brain maturation. *Brain connectivity*, *3*(1), 50-60.

Boersma, P. & Weenink, D. (2001). Praat, a system for doing phonetics by computer. Glot. Int., 5(9), 341-345. Computer program. Version. 6.0.37.

Bornstein, M.H., Hahn, C.S., & Haynes, O.M. (2011). Maternal personality, parenting cognitions, and parenting practices. Developmental Psychology, 47(3), 658-75.

Bornstein, M.H., Putnick, D.L., Bohr, Y., Abdelmaseh, M., Yookyung Lee, C., & Esposito, G. (2020). Maternal sensitivity and language in infancy each promotes child core language skill in preschool. Early Childhood Research Quarterly, 51, 483-489.

Breiman, L. (2001) Random Forests. *Machine Learning* 45, 5–32. https://doi.org/10.1023/A:1010933404324

Chen X, Ishwaran H. (2012) Random forests for genomic data analysis. *Genomics*. 99(6):323-9. doi: 10.1016/j.ygeno.2012.04.003.

Degenhardt, F., Seifert, S., Szymczak, S. (2019) Evaluation of variable selection methods for random forests and omics data sets, Briefings in Bioinformatics, Volume 20, Issue 2, Pages 492–503, https://doi.org/10.1093/bib/bbx124

Dehaene-Lambertz, G., Dehaene, S., & Hertz-Pannier, L. (2002). Functional neuroimaging of speech perception in infants. Science, 298(5600), 2013-2015.

Dehaene-Lambertz, G., Hertz-Pannier, L., Dubois, J., Mériaux, S., Roche, A., Sigman, M., & Dehaene, S. (2006). Functional organization of perisylvian activation during presentation of sentences in preverbal infants. Proceedings of the National Academy of Sciences, 103(38), 14240-14245.

Delorme, A., & Makeig, S. (2004). EEGLAB: an open-source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of neuroscience methods*, *134*(1), 9-21.

Dubois, J., Poupon, C., Thirion, B., Simonnet, H., Kulikova, S., & Leroy, F. & Dehaene-Lambertz, G.(2016). Exploring the early organization and maturation of linguistic pathways in the human infant brain. Cerebral Cortex, 26(5), 2283-2298.

Farkas, D., & Orosz, G. (2013). The link between ego-resiliency and changes in Big Five traits after decision making: The case of Extraversion. Personality and Individual Differences, 55, 440–445.

Gervain, J., Macagno, F., Cogoi, S., Peña, M., & Mehler, J. (2008). The neonate brain detects speech structure. Proceedings of the National Academy of Sciences, 105(37), 14222-14227.

Giraud, A. L., & Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. Nature neuroscience, 15(4), 511.

Greenwell B (2021). fastshap: Fast Approximate Shapley Values. R package version 0.0.7, https://cran.r-project.org/package=fastshap

Hartman, K.M., Ratner, N.B., & Newman, R.S. (2017). Infant-directed speech (IDS) vowel clarity and child language outcomes. Journal of Child Language, 44, 1140–1162.

Hastie, T., Tibshirani, R., Friedman, J. (2009). The elements of statistical learning: data mining, inference and prediction. Springer.

Hilton, C. B., Moser, C. J., Bertolo, M., Lee-Rubin, H., Amir, D., Bainbridge, C. M., ... & Mehr, S. A. (2022). Acoustic regularities in infant-directed speech and song across cultures. Nature Human Behaviour, 6(11), 1545-1556.

Kalashnikova, M., & Burnham, D. (2018). Infant-directed speech from seven to nineteen months has similar acoustic properties but different functions. Journal of Child Language, 45(5), 1035-1053.

Kisler, T., Reichel, U., & Schiel, F. (2017). Multilingual processing of speech via web services. Computer Speech & Language, 45, 326-347.

Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3? Perception 36 ECVP Abstract Supplement. *PLOS ONE*.

Kruskal, J.B. (1956). On the Shortest Spanning Subtree of a Graph and the Traveling Salesman Problem. Proceedings of the American Mathematical Society, 7, 48-50.

Kursa, M. B., & Rudnicki, W. R. (2010). Feature Selection with the Boruta Package. Journal of Statistical Software, 36(11), 1–13. https://doi.org/10.18637/jss.v036.i11

Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. Nature reviews neuroscience, 5(11), 831-843.

Leong, V., & Goswami, U. (2014). Assessment of rhythmic entrainment at multiple timescales in dyslexia: evidence for disruption to syllable timing. Hearing research, 308, 141-161.

Leong, V., & Goswami, U. (2015). Acoustic-emergent phonology in the amplitude envelope of child-directed speech. PloS one, 10(12), e0144411.

Leroy, F., Glasel, H., Dubois, J., Hertz-Pannier, L., Thirion, B., Mangin, J. F., & Dehaene-Lambertz, G. (2011). Early maturation of the linguistic dorsal pathway in human infants. Journal of Neuroscience, 31(4), 1500-1506.

Lundberg, S.M., & Lee, S. (2017). A Unified Approach to Interpreting Model Predictions. *NIPS*.

Madigan, S. L., Prime, H., Graham, S. A., Rodrigues, M., Anderson, N., Khoury, J. E., & Jenkins, J.M. (2019). Parenting Behavior and Child Language: A Meta-analysis. "Pediatrics", 144(4), 1-14.

Mády, K., Gyuris, B., Gärtner, H.-M., Kohári, A., Szalontai, Á., & Reichel, U. D. (2022). Perceived emotions in infant-directed narrative across time and speech acts. 11th International Conference on Speech Prosody. 23-26 May 2022. Lisbon.

Mayer M (2021). missRanger: Fast Imputation of Missing Values. R package version 2.1.3, https://cran.r-project.org/package=missRanger

McDonald, N. M., Perdue, K. L., Eilbott, J., Loyal, J., Shic, F., & Pelphrey, K. A. (2019). Infant brain responses to social sounds: A longitudinal functional near-infrared spectroscopy study. Developmental cognitive neuroscience, 36, 100638.

Prinzie, P., Stams, G. J. J., Deković, M., Reijntjes, A. H., & Belsky, J. (2009). The relations between parents' Big Five personality factors and parenting: A meta-analytic review. Journal of personality and social psychology, 97(2), 351–362.

Probst, P., Wright, M.N., Boulesteix, A.L. (2019) Hyperparameters and tuning strategies for random forest. Wiley Interdiscip. Rev. Data Mining Knowl. Discov. 2019, 9, e1301.

R Core Team (2022). R: A Language and Environmentfor Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Reichel, U. D. (2016). Copasul manual–contour-based parametric and superpositional intonation stylization. arXiv preprint arXiv:1612.04765.

Reilly, S., Wake, M., Bavin, E.L., Prior, M., Williams, J., Bretherton, L., Eadie, P., Barrett, Y., & Ukoumunne, O.C. (2007). Predicting Language at 2 Years of Age: A Prospective Community Study. Pediatrics 120 (6): e1441–e1449.

Robson, J., & Beck, J. M. (1999). Hearing smiles-Perceptual, acoustic and production aspects of labial spreading. In XIVth Proceedings of the XIVth International Congress of Phonetic Sciences. Volume 1: 219-222. International Congress of Phonetic Sciences.

Saint-Georges, C., Chetouani, M., Cassel, R., Apicella, F., Mahdhaoui, A., Muratori, F., Laznik, M.-C., & Cohen, D. (2013). Motherese in interaction: at the cross-road of emotion and cognition? (A systematic review). PloS one, 8(10), e78103.

Stam, C. J., Nolte, G., & Daffertshofer, A. (2007). Phase lag index: assessment of functional connectivity from multi-channel EEG and MEG with diminished bias from common sources. *Human brain mapping*, *28*(11), 1178-1193.

Stam, C. V., & Van Straaten, E. C. W. (2012). The organization of physiological brain networks. *Clinical neurophysiology*, *123*(6), 1067-1087.

Stam, C. J., Tewarie, P., Van Dellen, E., Van Straaten, E. C. W., Hillebrand, A., & Van Mieghem, P. (2014). The trees and the forest: characterization of complex brain networks with minimum spanning trees. *International Journal of Psychophysiology*, *92*(3), 129-138.

Stekhoven, D. J., Bühlmann, P. (2012) MissForest—non-parametric missing value imputation for mixed-type data, Bioinformatics, Volume 28, Issue 1, Pages 112–118, https://doi.org/10.1093/bioinformatics/btr597

Tewarie, P., van Dellen, E., Hillebrand, A., & Stam, C. J. (2015). The minimum spanning tree: an unbiased method for brain network analysis. *Neuroimage*, *104*, 177-188.

Tóth, B., Urbán, G., Haden, G. P., Márk, M., Török, M., Stam, C. J., & Winkler, I. (2017). Large-scale network organization of EEG functional connectivity in newborn infants. *Human Brain Mapping*, *38*(8), 4019-4033.

von Hippel, P. T. (2007). 4. Regression with Missing Ys: An Improved Strategy for Analyzing Multiply Imputed Data. Sociological Methodology, 37(1), 83–117. https://doi.org/10.1111/j.1467-9531.2007.00180.x

Waljee, A. K., Mukherjee, A., Singal, A. G., Zhang, Y., Warren, J., Balis, U., Marrero, J., Zhu, J., Higgins, P. (2013) Comparison of imputation methods for missing laboratory data in medicine, BMJ Open 3:e002847. doi: 10.1136/bmjopen-2013-002847

Wright, M. N. & Ziegler, A. (2017). ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. J Stat Softw 77:1-17. doi: 10.18637/jss.v077.i01

Zhao, T., Xu, Y., & He, Y. (2019). Graph theoretical modeling of baby brain networks. NeuroImage, 185, 711-727.