



Universiteit Utrecht

De invloed van de frequentie van zware lettergrepen op de leerbaarheid in domeinalgemene leerders van bepaalde kwantiteitssensitieve klemtoonsystemen

Auteur:
A.J.H. ZUIJDERWIJK

Begeleider:
dr. A.I. NAZAROV

Tweede beoordelaar:
prof. dr. H.E. de Swart

Een thesis van 7,5 ECTS voor het behalen van de graad van Bachelor of Science

in

Kunstmatige intelligentie

aan de

Faculteit Geesteswetenschappen

Juli 2020

Voorwoord

Voordat ik aan mijn studie begon had ik niet verwacht dat ik, een student kunstmatige intelligentie, me zou gaan verdiepen in een onderwerp waarin de taalkunde zo centraal zou staan als in dit onderzoek. Het multidisciplinaire karakter van dit onderzoek, *taalkunde meets informatica*, maakte dat ik nu een keer de algoritmes als middel kon gebruiken om een onderzoek te doen, in plaats van me te beperken tot onderzoek van álleen die algoritmes an sich. Dit onderzoek heeft mijn interessegebied echt uitgebreid.

Ik wil mijn begeleider, Aleksei Nazarov, zeer bedanken voor alle Teams-sessies, alle aanwijzingen en uitleg over velerlei taalkundige onderwerpen, maar bovenal zijn vermogen om me telkens met hernieuwde motivatie uit een sessie te doen komen. Daarnaast bood hij mij een exclusief kijkje achter de schermen bij zijn eigen wetenschappelijke onderzoek. Het was een zeer plezierige en vruchtbare samenwerking.

Omdat de Universiteitsbibliotheek vanwege de coronamaatregelen gesloten was, heb ik om de prikkelrijke omgeving van mijn eigen huis te ontvluchten zo nu en dan gebruikgemaakt van één van de ruimtes van de A.U.S.R. Orca, waarvoor (driewerf) dank.

Dan rest mij niets anders dan de lezer plezier te wensen met het doornemen van deze scriptie.

Jos Zuiderwijk

UNIVERSITEIT UTRECHT

*Samenvatting***De invloed van de frequentie van zware lettergrepen op de leerbaarheid in domeinalgemene leerders van bepaalde kwantiteitssensitieve klemtoonsystemen**

door A.J.H. ZUIJDERWIJK

De Expectation Driven Parameter Learner (EDPL) is een leerder van klemtoonsystemen die in 96% van de gevallen succesvol convergeert. Deze klemtoonsystemen worden geformaliseerd in het Chomskiaanse P&P-model en worden daarmee gepresenteerd als parameters. De systemen die niet geleerd werden, zgn. 'problematische talen', hebben bepaalde parametrische eigenschappen die verband houden met zware lettergrepen. In dit onderzoek onderzoek ik de invloed van de frequentie van zware lettergrepen op de leerbaarheid van die problematische talen. Dat doe ik door EDPL voor deze talen te laten runnen op datasets die variëren in lettergreepdistributie. Daaruit blijkt dat de talen niet meer problematisch zijn voor datasets waarin zware lettergrepen frequenter voorkomen; sterker nog: ze convergeren sneller op deze datasets dan talen zonder de oorspronkelijk problematische, parametrische eigenschappen. Laatstgenoemde groep convergeert sneller op de datasets die geen hogere frequentie aan zware lettergrepen hebben.

Inhoudsopgave

1	Introductie	1
2	Achtergrond	3
2.1	Lettergrepen	3
2.2	Representatie van de lettergreepstructuur	4
2.2.1	Onset-rhyme	4
2.2.2	Moraïsche theorie	5
2.3	Metrische voeten	5
2.4	Gewicht en beklemtoning	6
2.5	Principles & Parameters	7
2.6	Leerbaarheid van problematische talen	8
2.7	Problematische, natuurlijke talen	9
3	Methode	11
3.1	Model	11
3.2	Algoritme	11
3.2.1	EDPL	12
3.3	Datasets	14
3.4	Grammatica's	15
4	Resultaten	16
4.1	Dataset met lichte lettergrepen	16
5	Discussie	18
5.1	Leerbaarheid	18
5.2	Snelheid	18
6	Conclusie	19
	Referenties	21
A	Parameters	23
B	Volledige resultaten	25
B.1	Voorlopige resultaten D_3	29

1 Introductie

Menselijke taalverwerving is een complex fenomeen. Er is veel onderzoek gedaan naar hoe de mens zich van kinds af aan een taal eigen maakt. Chomsky poneerde de zogenaamde **Universele Grammatica** (UG), zijnde een aangeboren menselijk vermogen om taal te leren. Dit veronderstelt universele patronen in alle menselijke talen. Zulke patronen kunnen worden gerepresenteerd middels **Principles & Parameters** (P&P; Chomsky 1981/1982). In dat systeem zijn principes vaststaande regels, e.g. “elke zin heeft een onderwerp”, en parameters opties die waar danwel onwaar zijn, vergelijkbaar met schakelaars die je aan of uit kunt zetten. Op die manier kun je syntactische regels vatten in een verzameling principes en parameters.

Dat geldt ook voor klemtoonsystemen. Pearl (2011) poogt met verschillende, onbevooroordeelde, probabilistische leerders de juiste parametrische representatie voor het klemtoonpatroon van het Engels te vinden om taalacquisitie bij kinderen na te bootsen. Dat doet zij door woorden die kinderen frequent tegenkomen te nemen als input voor twee leerders: de **Naive Parameter Learner** (NPL; Yang 2002) met een lineair beloningssysteem en een leerder die Bayes' regel gebruikt om de kanswaarden te updaten. Zij concludeert dat onbevooroordeelde, probabilistische leerders niet in staat zijn om de correcte grammatica's te verwerven. Er zou bovenop P&P een extra laag aan informatie moeten worden toegevoegd die ervoor moet zorgen dat de leerder convergeert, bijvoorbeeld met een selective learning bias; de leerder is dan niet meer onbevooroordeeld.

Nazarov en Jarosz (2017) bestrijden Pearl door met een nieuw voorstel te komen voor een probabilistische leerder die klemtoonparameters verwerft. Zij bouwen voort op het algoritme van Yang met inzichten van Jarosz (2015) en stellen een nieuwe domeinalgemene, statistische leerder voor: de **Expectation Driven Parameter Learner** (EDPL), die een aantal problemen die Pearl beschrijft oplost. In § 3.2.1 worden die problemen verder toegelicht. EDPL convergeert in 96% van de gevallen met succes tegenover Pearls resultaat van slechts 4,3% convergentie bij het leren van de klemtoonparameters van Dresher en Kaye (1990) in 23 gegenereerde talen.

De hierboven genoemde EDPL faalt dus in 4% van de gevallen. Dat was volgens Nazarov en Jarosz het geval bij “een kwantiteitssensitieve L-naar-R trocheïsche taal met linker extrametricaliteit”.¹ Vervolgonderzoek (Nazarov en Jarosz 2020) wees uit dat andere kwantiteitssensitieve talen ook een probleem opleverde voor de leerder, onder andere talen waarin het type zware lettergreep een invloed had op de leerbaarheid. In talen met die specifieke parametrische eigenschappen komen zware lettergrepen vaker voor. Dat werpt de vraag op waarmee het falen van EDPL verband houdt.

In dit paper toets ik de hypothese dat de leerbaarheid in domeinalgemene leerders van de kwantiteitssensitieve klemtoonsystemen die door Nazarov en Jarosz (2017) niet geleerd werden, afhangt van de frequentie van zware lettergrepen.² Ten eerste zal ik de rol van de zware lettergreep bestuderen. Vervolgens zal ik bepalen welk bereik van distributies van lettergreepgewichten een domeinalgemene leerder succesvol kan laten convergeren. Als de hypothese klopt, vervalt het bezwaar van Pearl tegen onbevooroordeelde, domeinalgemene leerders die klemtoonsystemen leren.

¹Dit is dus een specifieke instantiatie van vier parameters uit Dresher en Kaye (1990).

²Apoussidou (2007, pp. 106–107) wijst ook in een vergelijkbaar geval naar de mogelijke invloed van een meerderheid aan zware lettergrepen op de leerbaarheid.

In het eerste hoofdstuk wordt de theoretische achtergrond behandeld. Daarbij ligt de nadruk op het lettergreepgewicht en de parameters. In het daaropvolgende hoofdstuk worden de algoritmes achter EDPL onder de loep genomen en worden de datasets nader bestudeerd. Vervolgens wordt een experiment met meerdere distributies van zware lettergrepen als input aan EDPL toegelicht. In de hoofdstukken daarna worden de resultaten en de implicaties daarvan besproken.

Het experiment zal bestaan uit het runnen van EDPL op verschillende datasets. Ik gebruik datasets die variëren in distributie van zware lettergreep en in het type van de zware lettergreep, i.e., zwaar in de nucleus of ook in de coda. Vervolgens hebben we originele dataset uit Nazarov en Jarosz (2017) ter controle. We nemen vier instantiaties van parameters (grammatica's) waarmee we kunnen toetsen in hoeverre de leerbaarheid afhangt van de distributie van de zware lettergrepen.

Op deze manier boots ik menselijke taalverwerving machinaal na. De kern van het onderzoek ligt in de vergelijking tussen enerzijds de verschillende datasets die veel of weinig zware lettergrepen bevatten en anderzijds de verschillende grammatica's. Daarvan zijn twee problematisch volgens Nazarov en Jarosz (2017) en twee zonder problematische parameter, zie ook § 3.4.

Relevantie binnen kunstmatige intelligentie

Theoretische taalkunde is de stroming binnen de formele taalkunde die grammaticale verschijnselen in systematische modellen vat. Chomsky (1957) was de grondlegger van de generatieve taalkunde, een deelgebied waarin de hypothese van de universele grammatica centraal staat. Daarmee kreeg de taalkunde raakvlakken met de cognitieve psychologie.

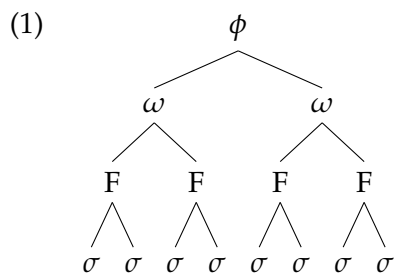
Machinale leerders kunnen op basis van deze theorie onder andere de menselijke taalverwerving nabootsen. Kunstmatige intelligentie is daarbij niet alleen een doel op zich; het kan in dit geval inzicht bieden in welke psychologische mechanismen er beschikbaar zijn bij het leren van taal bij mensen. Dat zou een bijdrage kunnen leveren in onderzoek naar taalverwerving.

2 Achtergrond

2.1 Lettergrepen

Een lettergreep of syllabe (σ) is een belangrijk begrip in de taalkunde (e.g. Ewen en Van der Hulst 2001). In deze scriptie zal ik niet uitgebreid op de conceptuele achtergrond ingaan, maar wel de theorieën passend bij dit onderzoek aanstippen.

Ewen en Van der Hulst (2001) beschrijven een lettergreep als een prosodische eenheid. De lettergreep kan worden gezien als onderdeel van een hiërarchische structuur die wordt aangeduid als de **prosodische hiërarchie**. Deze wordt gevisualiseerd in (1). We geven ϕ weer als fonologische frase, ω als fonologisch woord, F als voet en σ als lettergreep.



In dit paper zal ik niet uitwiden over de fonologische frase en het fonologische woord. De voet zal ik in bespreken in § 2.3.

Ewen en Van der Hulst wijzen erop dat de prosodische hiërarchie, waarin iedere constituent zich splitst in twee lagere constituenten, een afspiegeling is van het principe geformuleerd door Nespor en Vogel (1986) als: "A given nonterminal unit of the prosodic hierarchy, X^p , is composed of one or more units of the immediately lower category X^{p-1} ". Een voet bestaat uit twee lettergrepen en de lettergreep is dus de kleinste eenheid. In latere theorieën (McCarthy en Prince 1993) wordt de mora (zie § 2.2.2) ook beschouwd als onderdeel van de prosodische hiërarchie, namelijk onder de lettergrepen.

Moedertaalsprekers zijn in staat te bepalen hoeveel en welke lettergrepen zich in een woord bevinden. Voorbeeld (2) uit Ewen en Van der Hulst (2001, p. 120) laat zien hoe de meeste sprekers van het Engels de volgende woorden zouden syllabificeren.

- (2) albatross [æɪ]_σ[bə]_σ[trɒs]_σ
 America [ə]_σ[mɛ]_σ[rɪ]_σ[kə]_σ
 slender [slɛn]_σ[də]_σ

Dat wil niet zeggen dat het altijd zo simpel is als het vorige voorbeeld suggereert. Zo zien we in (3) dat sprekers niet altijd op één lijn zitten bij het plaatsen van de lettergreepgrenzen.

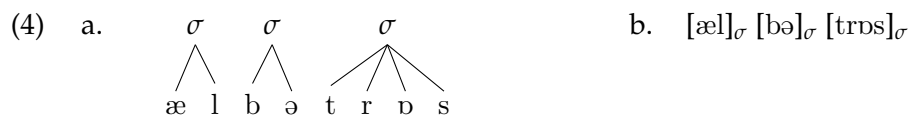
- (3) master [ma:]_σ[stə]_σ of [ma:s]_σ[tə]_σ
 revels [rɛ]_σ[vəlz]_σ of [rɛv]_σ[əlz]_σ
 pastry [peɪ]_σ[stri]_σ of [peɪs]_σ[tri]_σ of [peɪst]_σ[ri]_σ

Ondanks de verschillen in (3) mogen we aannemen dat er een eenheid is die sprekers herkennen als een lettergreep. De lettergreepanalyse baseert zich dan ook vaak op verdere factoren, zoals fonotaxis (mogelijke woorden/klankcombinaties), (klem)toon en fonologische processen. Een voorbeeld van Ewen en Van der Hulst (2001) uit de fonotaxis is het feit dat moedertaalsprekers kunnen bepalen of een woord 'mogelijk' danwel 'onmogelijk' is in de betreffende taal. Een Nederlander zou het woord

*/lmɔk/ resoluut afwijzen, omdat een woord in het Nederlands nooit met /lm/ kan beginnen, i.e., dat schendt de *fonotactische beperkingen* van het Nederlands. Vanuit de fonotaxis is het mogelijk om lettergreepstructuren te ontwaren.

2.2 Representatie van de lettergreepstructuur

Er zijn in de loop der jaren verschillende theoriën geponeerd die de structuur van lettergrepen beschrijven. De minimalistische representatie van het Engelse *albatross*, gebaseerd op het algemene, formele idee van een boomstructuur, zou eruit kunnen zien als in (4).



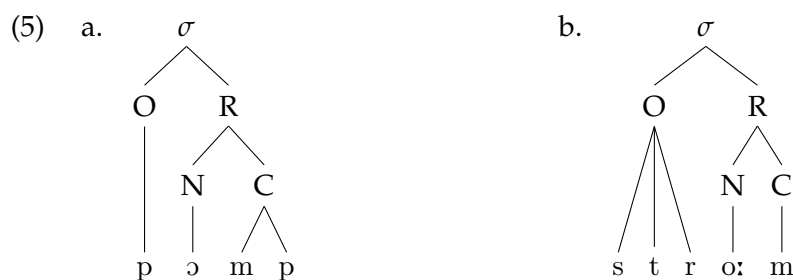
Merk op dat (4a) en (4b) equivalent zijn. De lijnen in (4a) duiden op een hiërarchische structuur, i.e., /æ/ en /l/ zijn onderdeel van een eenheid σ . (4b) is een notationale variant.

Het voorbeeld uit (4) is plat, zonder interne structuur. Er zijn meerdere voorstellen gedaan waarin zo'n interne structuur, die je kunt voorstellen als één of meerdere extra lagen knooppunten tussen de wortel en de onderste punten, wel aanwezig is. Eerst beschouwen we de **onset-rhyme theorie**. Vervolgens kijken we naar de **moraische theorie**.

2.2.1 Onset-rhyme

Een lettergreep is opgebouwd uit verschillende componenten: de **nucleus**, of de kern (N) bestaat uit één of twee vocalen. Deze wordt eventueel voorafgegaan door een **aanzet** (O) en eventueel gevolgd door een **coda** (C). De aanzet en de coda bestaan uit één of meerdere consonanten. De nucleus en de coda worden gegroepeerd onder de noemer rijm (R).

Neem de monosyllabische voorbeelden /pɔmp/ (pomp) en /stro:m/ (stroom).



We zien in voorbeeld (5) dat de klinker onderdeel is van de nucleus. In een voorbeeld als /'oma/ (oma) hebben we te maken met twee lettergrepen: [o] σ en [ma] σ . De eerste lettergreep bestaat dan slechts uit een nucleus. De tweede lettergreep bestaat uit een nucleus en een aanzet.

Met 'C' en 'V' als afkortingen voor 'consonant' respectievelijk 'vocaal' kunnen we generaliserend over lettergrepen spreken. Woorden als *kat*, *lik* en *pot* zijn van de vorm CVC. Voorbeeld (5a) is van de vorm CVCC. Lange vocalen en diftongen worden onderscheiden van korte vocalen (V) door ze te noteren als VV. Voorbeeld (5b) wordt dan CCCVVC.

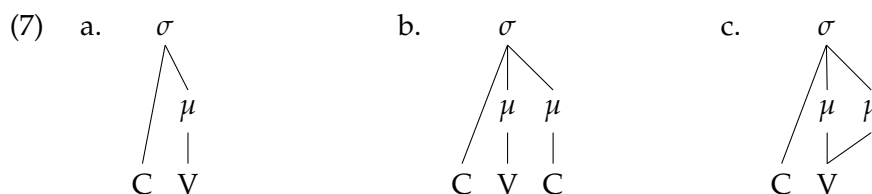
2.2.2 Moraïsche theorie

De moraïsche theorie gaat uit van zogenaamde **morae** (afkomstig uit het Latijn, met als betekenis ‘vertraging’). Ryan (2016) stelt dat sinds de late jaren 1980 de meeste taalkundigen de mora (μ) accepteren als maat van lettergreepgewicht.

Hayes (1989) vat het als volgt samen:³

- (6) a. Lichte lettergrepen bestaan uit één mora
b. Zware lettergrepen bestaan uit twee morae

Neem een taal waarin zware lettergrepen van de vorm CVC moeten zijn (een gesloten lettergreep) of een lange klinker moeten bevatten. Dan kunnen we de lettergreepstructuur als volgt beschrijven:



In die taal geldt (7a) als monomoraïsch c.q. licht, en (7b) en (7c) als bimoraïsch c.q. zwaar. De mora is in dit systeem dus een onderdeel van de lettergreep.

De moraïsche theorie werd als tegenhanger van de onset-rhyme theorie gepresenteerd voor het analyseren van fonologische processen. Tegenwoordig wordt bijna altijd impliciet een link aangenomen in de literatuur. De combinatie van de twee theoriën houdt dus in dat er zowel wordt gesproken van bijvoorbeeld coda’s én morae, e.g. Hayes (1989) die spreekt van *weight by position*, namelijk dat de codaconsonant een mora krijgt en de onsetconsonant niet.

Morae worden net als lettergrepen gezien als een prosodische eenheid, zie ook McCarthy en Prince (1993).

2.3 Metrische voeten

De (metrische) voet is een prosodische eenheid die direct boven de lettergreep staat (zie 1). Lettergrepen worden dus gegroepeerd door voeten. Twee simpele voorbeelden van een voet zijn de **jambe** en de **trochee**. Deze twee metrische voeten, beide bestaande uit een paar van lettergrepen, alterneren tussen beklemtoonde en onbeklemtoonde lettergrepen. In een trochaïsche voet is de eerste lettergreep beklemtoond en de tweede onbeklemtoond; een jambische voet laat het tegenovergestelde patroon zien (Ewen en Van der Hulst 2001).

- (8) Jambe ($\sigma . \acute{\sigma}$) ‘trochée’
Trochee ($\acute{\sigma} . \sigma$) ‘jámbe’

Ironisch genoeg heeft het woord “jambe” een trochaïsche structuur en het woord “trochee” een jambische structuur. In het onderzochte framework in deze scriptie komen we alleen jambische en trochaïsche voeten tegen.

De facto geldt de voet als eenheid van klemtonen, aangezien iedere klemtoon een hoofd van een voet moet zijn. Merk wel op dat niet elk hoofd van een voet als

³Daarbij veronderstelt Hayes ook lettergrepen bestaande uit drie morae, zogenaamde superzware lettergrepen, die in sommige talen zouden voorkomen. Die laten we buiten beschouwing.

klemtoon wordt gerealiseerd; iedere klemtoon is een hoofd van een voet, maar niet per definitie vice versa. Op dezelfde manier kan het prosodische woord worden gezien als eenheid van woordklemtonen, omdat de hoofdvoet de woordklemtoon (hoofdklemtoon) levert.

Vanuit die invalshoek, kunnen de Latijnse klemtoonregels uit (9) ook beschreven worden in termen van voeten. Daarvan geef ik aan het einde van § 2.4 een uitwerking.

2.4 Gewicht en beklemtoning

Zoals we in § 2.2.2 hebben gezien, kan een lettergreep dus een bepaald gewicht hebben, uitgedrukt in μ . Een lettergreep is ofwel licht, ofwel zwaar, zoals samengevat in (6). Specifieker gezegd hangt het gewicht van een lettergreep in veruit de meeste talen⁴ af van de rijm, dat wil zeggen een optelsom van de morae uit nucleus en coda.

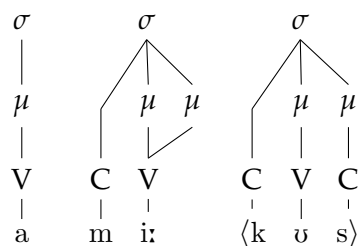
Toch is het niet absoluut te duiden of een lettergreep zwaar danwel licht is; dit is erg taalafhankelijk. Neem een taal als Latijn, die de volgende, relatief eenvoudige klemtoonregels heeft (Hayes 1995):

- (9)
- Latijn kent zowel korte als lange klinkers.
 - De klemtoon ligt op de voorlaatste (penult) of de voorvoorlaatste (antepenult) lettergreep.
 - De penult wordt beklemtoond als die lettergreep lang is.
 - De antepenult wordt beklemtoond als de penult kort is, ongeacht of de antepenult zelf lang of kort is.

Latijn is een **kwantiteitssensitieve** taal. Dit begrip, geïntroduceerd door Hayes (1981) is één van de parameters (P5) die betrekking hebben op een klemtoonsysteem. Een taal (c.q. klemtoonsysteem) is kwantiteitssensitief als de voeten afhankelijk zijn van de interne structuur van de lettergreep, i.e. het hoofd van een voet is altijd een zware lettergreep. Bij kwantiteitsinsensitieve talen kan het hoofd van een voet zowel licht als zwaar zijn. Drescher en Kaye (1990) noemen dat je kwantiteitssensitiviteit kunt herkennen in een taal als woorden met hetzelfde aantal lettergrepen een ander klemtoonpatroon hebben, e.g. *símula* en *amícus*. Zo'n observatie waaraan je een parameter kunt herkennen heet een *cue* (Drescher en Kaye 1990).

Voorts is Latijn **rechts-extrametricaal**; van rechts naar links gezien wordt de eerste syllabe niet meegerekend in een metrische voet. Hayes (1995, p. 91) geeft drie voorbeelden van verschillende type woorden. Een korte klinker noteren we met een breve, e.g. *ă* en een lange klinker met een macron, e.g. *ā*.

(10) *ămīcūs*



⁴Een tegenvoorbeeld is Pirahã, waarbij de aanzet lijkt mee te tellen voor het gewicht (Van Oostendorp 2005)

- P3: Feet are built from the [Left/Right] (DIR=L-TO-R/R-TO-L)
- P4: Feet are strong on the [Left/Right] (FOOT=TROCHEE/IAMB)
- P5: Feet are quantity sensitive (QS) [Yes/No] (QS=ON/OFF)
- P6: Feet are QS to the [Rime/Nucleus] (CVC=LIGHT/HEAVY)
- P7: A strong branch of a foot must itself branch [No/Yes] (HEAVYHEAD=ON/OFF)
- P8A: There is an extrametrical syllable [No/Yes] (XM=ON/OFF)
- P8: It is extrametrical on the [Left/Right] (XMDIR=LEFT/RIGHT)
- P9: Feet consisting of a single light syllable are removed [No/Yes] (DEGENERATE=ON/OFF)
- P10: Feet are noniterative [No/Yes] (SECSTRESS=ON/OFF)

Enkele van deze parameters (P5, P8) zijn al beschreven in § 2.4. In Appendix A staat het restant van de parameters beschreven.

2.6 Leerbaarheid van problematische talen

Klemtoonsystemen, die gerepresenteerd worden door middel van een verzameling van elf parameters, worden met de aanname van Chomsky's **Universele Grammatica (UG)** verondersteld als een aangeboren menselijke basis voor het uitspreken van klemtonen. In de basis komt het erop neer dat de mens een achterliggend vermogen heeft om taal(structuren) te kunnen leren.

Pearl (2011) probeert om met computationele leerders zulke klemtoonsystemen te laten convergeren naar de Engelse grammatica. Om de menselijke taalverwerving na te bootsen, maakt Pearl gebruik van een dataset met woorden die kinderen het vaakst tegenkomen. Het blijkt dat de NPL zeer slecht in staat is om te convergeren naar de juiste grammatica's. De **Naive Parameter Learner** (NPL; Yang 2002) is slechts in 4,3% van de gevallen succesvol. De conclusie is dat onbevooroordeelde, probabalistische leerders niet in staat zijn om de klemtoonsystemen te leren.

Vervolgens oppert Pearl dat met een extra domeinspecifieke leermechanismen bovenop P&P het wel mogelijk zou moeten zijn voor de leerder, e.g. met een *bevooroordeelde* (biased) leerder. Eerder werk (Pearl 2008, 2009) toont aan dat een selectieve learning bias succesvol bleek. Er werd alleen geleerd van de beschikbare data die werd gezien als maximaal informatief: ondubbelzinnige data (Fodor 1998; Dresher 1999; Lightfoot 1999; Pearl en Weinberg 2007). In het geval van systemen met parameters die elkaar beïnvloeden (zoals klemtoonsystemen), is die ondubbelzinnige data slechts een kleine deelverzameling. Bovendien is het niet eenvoudig om ondubbelzinnige data aan te wijzen, al is daar wel onderzoek naar gedaan.⁵ Naast de selectieve bias, moeten de leerders volgens Pearl (2008, 2009) een bepaalde prioritering van parameters handhaven. Leerders die aan die voorwaarden voldoen zijn wel succesvol. In conclusie zou het dan aan de onbevooroordeeldheid van de leerders te wijten zijn dat zij niet succesvol zijn. Dit heeft volgens Pearl (2011) ook implicaties voor de manier waarop kinderen zich het Engelse klemtoonsysteem eigen maken; kinderen moeten dan biases hebben als ze hun kennis van de metrische fonologie op de manier representeren zoals dit parametrische systeem dat veronderstelt.

Nazarov en Jarosz (2017) komen met een nieuwe probabalistische leerder, gebaseerd op de eerdergenoemde NPL met inzichten van Jarosz (2015). Deze leerder krijgt de naam **Expectation Driven Parameter Learner (EDPL)** en convergeert in

⁵e.g. Dresher (1999), die stelt dat kinderen naar zgn. *cues* in de data zoeken.

96% van de gevallen. De talen waarmee EDPL moeite lijkt te hebben, worden gekenmerkt door veel ambiguïteit, o.a. veroorzaakt door kwantiteitssensitiviteit (QS=ON) en noniterativiteit (SECSTRESS = ON). Deze talen bevatten relatief veel zware lettergrepen. Nazarov en Jarosz (2020) vermeldt in vervolgonderzoek dat EDPL eveneens moeite heeft met andere kwantiteitssensitieve talen, waaronder die met CVC=HEAVY.

Het probleem is dat er natuurlijke talen bestaan die de gevonden problematische grammatica's volgen, e.g. Frans en Polabisch. In § 2.7 geef ik een voorbeelduitwerking van het Frans. Om de bewering te kunnen steunen dat onbevooroordeelde, probabilistische leerders in staat zijn om klemtoonssystemen te leren, moeten talen als Frans ook succesvol geleerd kunnen worden; het kan in dat geval dus geen inherent probleem van de problematische parameters zijn.

Nu hoeft de correlatie tussen de problematische parameters niet te duiden op causaliteit; een mogelijke oorzaak lijkt de frequentie van zware lettergrepen binnen deze grammatica's te zijn. In dit onderzoek vergelijk ik verschillende grammatica's, met als input verschillende distributies van zware lettergrepen. Hierover wijd ik uit in § 3.

2.7 Problematische, natuurlijke talen

Er is een aantal talen waarvan de sprekers de klemtoonpatronen volgen die beschreven worden door de problematische grammatica's (zie § 3.4). Met behulp van *StressTyp2* (ST2, Goedemans e.a. 2015) is het mogelijk om te zoeken naar talen waarvan het klemtoonpatroon voldoet aan bepaalde criteria. De talen die we zoeken moeten aan het patroon voldoen dat we zien in Tabel 1. Voor dit voorbeeld gebruik ik alvast de kunsttaal die nader toegelicht wordt in § 3.3.

MAIN (P1)	CVC (P6)	“tan-taa-ta-taa-tan”	“ta-ta-ta-ta-ta”
LEFT	LIGHT	(tan'taa)(ta.taa)tan	(ta'ta)(ta.ta)ta
LEFT	HEAVY	('tan)(taa)(ta.taa)(tan)	(ta'ta)(ta.ta)ta
RIGHT	LIGHT	tan(taa.ta)('taa.tan)	ta(ta.ta)('ta.ta)
RIGHT	HEAV	(tan)(taa.ta)(taa)('tan)	ta(ta.ta)('ta.ta)

TABEL 1: Twee voorbeelden van het klemtoonpatroon

We zoeken talen met *syllable priority code*: 12/2R⁶ in ST2. Daarnaast mag de taal geen nevenklemtonen bevatten (SECSTRESS=OFF) en CVC-lettergrepen bevatten die als licht tellen. De volgende talen werden gevonden:

- Frans
- Polabisch
- Bulisch
- Rotumaans
- Tobèlorees

Laten we eerst de talen doorspreken die alsnog afvallen. In het Bulisch komen lange klinkers namelijk alleen voor in beklemtoonde eindlettergrepen. De klemtoon ligt doorgaans op de penult, maar kán ook op een laatste lettergreep vallen met een korte klinker. (Maan 1953, pp. 13, 16).

Afhankelijk van hoe in het Rotumaans woorden eindigend op –CVC worden geanalyseerd, lijkt het erop dat CVC-lettergrepen niet beklemtoond worden.

⁶Deze code, ontwikkeld door Bailey (1995), betekent in dit geval: “als de eerste syllabe zwaar is, wordt die beklemtoond, anders: als de tweede syllabe zwaar is, wordt die beklemtoond. In alle andere gevallen: beklemtoon de tweede syllabe van rechts (antepenult).”

Het Tobèlorees heeft lange klinkers, maar CVC-lettergrepen zijn erg zeldzaam: die komen slechts voor aan het eind van leenwoorden. Het is ook niet duidelijk of deze lettergrepen dan licht zijn. De klemtoon valt op de antepenult als de laatste klinker epenthetisch⁷ is (Hueting 1936, pp. 324–325).

Dan houden we over het Frans en het Polabisch. De Polabische taal is een in de 18^e eeuw uitgestorven, Slavische taal waar we toch gedetailleerde informatie over hebben door een aantal teksten dat is geschreven vlak voordat de taal uitstierf (Stone 1972). Het Polabisch wordt gekenmerkt door een systeem waarin alle klinkers behalve de schwa-achtige, i.e., gereduceerde klinkers als zwaar tellen. De klemtoon komt op de laatste niet-gereduceerde klinker; in de praktijk valt de klemtoon dan op de voorlaatste klinker (als de laatste gereduceerd is) of de laatste (als deze niet-gereduceerd is). (Bethin 1998, pp. 180–182)

Selkirk (1978) stelt in haar analyse van het Frans dat alle vocalen lang zijn, met uitzondering van de /ə/ (schwa). Een Franse voet bestaat volgens Selkirk uit één lettergreep of uit een paar van lettergrepen waarvan de laatste een schwa is. Van der Hulst en Moortgat (1980, p. 15) geven de volgende voorbeelden.

- (13) a. (fo)(no)(lo)(‘zi)
 ‘phonologie’
 b. (bo)(tu)(‘lis.mœ)
 ‘botulisme’

Het Frans bevat dus alleen unaire en binaire voeten. In (13a) zien we een voorbeeld van een woord met alleen enkelvoudige voeten. In (13b) bestaat de laatste voet uit twee lettergrepen, waarbij de /œ/ zich als een schwa gedraagt. Het accent komt te liggen op het hoofd van die voet, i.e., de zware lettergreep /lis/.

In het Frans komt de klemtoon altijd op de laatste voet. Voeten worden van rechts naar links opgebouwd (Van der Hulst en Moortgat 1980). Daarnaast is het Frans kwantiteitssensitief; als we eerdergenoemde cue (Dresher en Kaye 1990, p. 190) nemen, zien we bewijs in (13) voor kwantiteitssensitiviteit: de beide voorbeelden hebben evenveel lettergrepen maar een ander klemtoonpatroon.

- (14) a.
$$\begin{array}{c} F \\ | \\ \sigma_Z \end{array}$$

 b.
$$\begin{array}{c} F \\ \swarrow \quad \searrow \\ \sigma_Z \quad \sigma \\ \quad \quad | \\ \quad \quad \text{ə} \end{array}$$

Van der Hulst en Moortgat (1980, p. 14) vatten Selkirks analyse samen in (14). De trochaische structuur is te herkennen in (14b). Uit deze analyses zijn de parameterwaarden te herleiden als volgt.

- (15) MAIN: RIGHT
 BOUNDED: ON
 DIR: R-TO-L
 FOOT: TROCHEE
 QS: ON
 CVC: HEAVY
 XM: OFF

Hiermee is het Frans vergelijkbaar met de problematische grammatica beschreven in § 3.4. De aanwijzingen beschreven om parameterwaarden te herkennen bij talen, zijn beschreven door Dresher en Kaye (1990, p.190–194).

⁷Epenthese is het inlassen van één of meerdere klinkers waarvoor geen etymologische reden is, vaak om de uitspraak te vergemakkelijken.

3 Methode

3.1 Model

De zoekruimte in dit model bestaat uit alle mogelijke combinaties van de parameters van Drescher en Kaye (1990). Aangezien dit binaire parameters betreft, zijn er $2^{11} = 2,048$ mogelijke grammatica's waarnaar de leerder kan convergeren. Niet iedere theoretisch mogelijke parametercombinatie leidt tot verschillende patronen. Nazarov en Jarosz (2020) reduceren dit getal met aanwijzingen van Drescher en Kaye (1990) (parameters zijn niet allemaal onafhankelijk, p. 147) tot 280 grammatica's. Niet iedere structuur heeft een unieke realisatie en niet iedere realisatie heeft een unieke structuur. We zouden de relatie tussen alle mogelijke combinaties en de deelverzameling van de zoekruimte dus kunnen zien als veel-op-veel.

Een bruteforcemethode zou misschien voor de hand liggen; zo'n kleine zoekruimte is daarvoor bij uitstek geschikt. Echter correspondeert die methode niet met de menselijke manier van taalverwerving. De literatuur (e.g. Fikkert 1994, Levelt e.a. 2000, Demuth 1996) wijst er eenduidig op dat de menselijke taalverwerving wordt gekarakteriseerd door kleine progressie tussen nauw verwante grammatica's, in tegenstelling tot het springen tussen ongerelateerde grammatica's waar bruteforce naar neigt. Het leeralgoritme moet dus bij voorkeur tussen verwante grammatica's schakelen totdat de grammatica met de kleinste foutmarge (of de grootste waarschijnlijkheid) gevonden wordt.

3.2 Algoritme

Een voorbeeld van zo'n algoritme is Yang (2002)'s NPL, een domeinalgemene, statistische leerder voor parametersystemen. Het door Nazarov en Jarosz (2017) voorgestelde EDPL is voor het grootste deel gebaseerd op de NPL; beide leerders delen hetzelfde model voor de parameters en dezelfde lineaire updateregels, die hierna besproken zullen worden. Het verschil ligt in de manier hoe de *reward/penalty* waarden worden uitgerekend, zie § 3.2.1.

Yang (2002) definieert een probabilistische parameter grammar (PPG) als een verzameling Bernoulli-distributies. Dat wil zeggen: iedere (binair) parameterwaarde uit (21) krijgt een waarschijnlijkheidswaarde. Onder parameterwaarde verstaan we een waarde die een parameter kan hebben, e.g. $\psi_i = \text{MAIN: LEFT}$. De negatie, we hebben immers binaire parameters, is dan $\neg\psi_i = \text{MAIN: RIGHT}$. Voor iedere parameter moet gelden $P(\psi_i) + P(\neg\psi_i) = 1$. De waarden van verschillende parameters onderling hebben geen verband. Ter illustratie een voorbeeld uit Nazarov en Jarosz (2020, p.11):

$$(16) \quad G_t = \left\{ \begin{array}{l} \text{MAIN: } P(\text{L}) = 0,6 \\ \text{ } P(\text{R}) = 0,4 \end{array} , \text{BOUNDED: } \begin{array}{l} P(\text{ON}) = 1 \\ P(\text{OFF}) = 0 \end{array} , \text{DIR: } \begin{array}{l} P(\text{L-TO-R}) = 0,3 \\ P(\text{R-TO-L}) = 0,7 \end{array} , \dots \right\}$$

Deze PPG wordt in dit geval gebruikt om de klemtoonwaarden te genereren voor een verzameling lettergrepen. Eerst krijgen alle parameters waarden op basis van de waarschijnlijkheidswaarden voor de betreffende parameter, een zogenaamde full parameter specification (FPS; Nazarov en Jarosz 2020). De output is gebaseerd op de FPS. Hieronder twee voorbeelden van mogelijke FPS's.

$$(17) \quad \text{a. FPS 1: MAIN = RIGHT, BOUNDED = ON, DIR = L-TO-R, ...}$$

$$P(\psi_1, \psi_2, \psi_3) = 0,4 \times 1 \times 0,3 = 0,12$$

Datapunt: ta-ta-ta-ta

Output: (,ta.ta)(,ta.ta)ta [10200]
 Correct: (,ta.ta)(,ta.ta)ta [10200] MATCH

b. FPS 2: MAIN = LEFT, BOUNDED = ON, DIR = R-TO-L, ...
 $P(\psi_1, \psi_2, \psi_3) = 0,6 \times 1 \times 0,7 = 0,42$
 Datapunt: ta-ta-ta-ta-ta
 Output: ta(,ta.ta)(,ta.ta) [02010]
 Correct: (,ta.ta)(,ta.ta)ta [10200] MISMATCH

In (17) worden twee FPS's gegenereerd uit de PPG uit 3.2. De kans op FPS 1 uit (17b) is groter dan op FPS 2 uit (17a), getuige de kanswaarden van de parameters. In FPS 1 geldt rechterhoofdbeklemtoning en worden voeten van links naar rechts opgebouwd. In FPS 2 zijn de waarden van deze parameters andersom, dat wil zeggen linkerhoofdbeklemtoning en voeten van rechts naar links. Merk op dat de voeten altijd gebonden moeten zijn, omdat de kans daarop in de PPG 100% is.

De leerder start met een onbevooroordeelde initialisatie, i.e., $\forall i P(\psi_i) = 0,5$. Vervolgens worden die waarden online geüpdatet als de datapunten één voor één verwerkt worden. Bij het leren wordt deze gegenereerde, verwachte output afgezet tegen een gegeven, correcte output. De parameterwaarden uit de PPG worden zowel in NPL als EDPL geüpdatet volgens de Linear Reward-Penalty Scheme (LRPS; Bush en Mosteller 1951):

(18) Voor iedere parameter ψ_i en iedere iteratie t :

$$\hat{P}(\psi_i | G_{t+1}) = \lambda \times R(\psi_i) + (1 - \lambda) \times P(\psi_i | G_t)$$

Waarbij:

- $P(\psi_i | G_t)$ de waarschijnlijkheid van de parameter in de PPG is op tijdstip t
- $R(\psi_i) \in [0, 1]$ de parameters huidige beloningswaarde is
- $\lambda \in [0, 1]$ de leergraad (*learning rate*) is; voor EDPL is dat $\lambda = 0,1$

Volgens deze formule is de waarschijnlijkheid van parameter ψ_i op tijdstip $t + 1$ een gewogen som van zijn waarschijnlijkheid op tijdstip t en zijn beloningswaarde (*reward value*; Nazarov en Jarosz 2017) genoteerd als $R(\psi_i)$. De leergraad λ , een waarde tussen 0 en 1, geeft aan in welke mate de update meeweegt: hoe lager λ , hoe minder de update meeweegt. De beloningswaarde heeft invloed op hoe de kans geüpdatet wordt. Bij $R(\psi_i) = 0$ zal de kanswaarde dalen met $\lambda \times \hat{P}(\psi_i | G_t)$, terwijl bij $R(\psi_i) = 1$ de kanswaarde juist met die hoeveelheid zal toenemen. Bij $R(\psi_i) \approx P(\psi_i | G_t)$ zal de waarschijnlijkheidswaarde ongeveer hetzelfde blijven.

3.2.1 EDPL

NPL en EDPL verschillen in de manier waarop $R(\psi_i)$ wordt berekend. Voor de NPL geldt dat iedere beloningswaarde 0 danwel 1 is, afhankelijk van één FPS die gesampled wordt van de huidige grammatica. De waarde voor $R(\psi_i) = 1$ als de waarde van die parameter uit het datapunt wordt gehaald ($\psi_i \in Current$) en er een match is, of andersom: als de tegenovergestelde waarde uit het datapunt wordt gehaald ($\neg\psi_i \in Current$) en er een mismatch is; anders $R(\psi_i) = 0$. Er wordt per datapunt dus één sample gebruikt zonder die te optimaliseren.

Yang (2002) erkent dat NPL geen oplossing biedt voor het *Credit probleem* (Clark 1990, 1992, zie ook Drescher 1999). Yang noemt twee manifestaties van dat probleem:

de lifter (*hitchhiker*) en de handlanger (*accomplice*). Kort gezegd treedt het lifersscenario op als een parameterwaarde beloond wordt terwijl de waarde in kwestie geen invloed had op de match. Informeel gezegd lift zo'n irrelevante parameterwaarde dan mee op het succes van een andere, relevante parameterwaarde. Het handlangersscenario lijkt hierop, maar werkt precies andersom. Dit scenario treedt op als een irrelevante parameterwaarde juist bestraft wordt, terwijl de mismatch veroorzaakt wordt door een andere, relevante parameterwaarde. Hierdoor komt het voor dat dezelfde parameter per ongeluk beloond wordt voor het ene datapunt, en per ongeluk gestraft wordt voor het volgende datapunt. Dit leidt tot substantiële ruis in NPL-leerders, zoals bijvoorbeeld die gebruikt door Pearl (2011).

EDPL lost dit op door middel van een geavanceerdere beloningsstrategie. EDPL stelt $R(\psi_i)$ gelijk aan de waarschijnlijkheid van parameterwaarde ψ_i gegeven het huidige datapunt X en de huidige grammatica G_t , zie (19b). Nazarov en Jarosz (2017) beschrijven deze strategie als een soort *Expectation Maximization* (Dempster e.a. 1977). Een EM-benadering zou zijn: alle kansen van parameterwaarden zo vaststellen dat ze overeenkomen met hun respectievelijke verwachte waarden gegeven de gehele dataset. EDPL (en NPL) is een online variant, waarbij de berekeningen plaatsvinden bij de verwerking van ieder nieuw datapunt.

$$(19) \quad \begin{aligned} \text{a. } \hat{P}(X | \psi_i, G_t) &= \frac{\text{aantal matches gegeven } G_t \text{ met } P(\psi_i)=1}{r} \\ \text{b. } R(\psi_i) \equiv \hat{P}(\psi_i | X, G_t) &= \frac{\hat{P}(X | \psi_i, G_t) \times P(\psi_i | G_t)}{\hat{P}(X | G_t)} \\ \text{c. } \hat{P}(X | G_t) &= \hat{P}(X | \psi_i, G_t) \times P(\psi_i | G_t) + \hat{P}(X | \neg\psi_i, G_t) \times P(\neg\psi_i | G_t) \end{aligned}$$

Lees $P(X)$ als "de kans op een match met datapunt X ". Om $\hat{P}(X | \psi_i, G_t)$ te benaderen, wordt er een tijdelijke grammatica G'_t gegenereerd, waarin de kans op parameterwaarde ψ_i wordt vervangen door 1 (en evenzo $P(\neg\psi_i) = 0$). Op deze manier is het zeker dat deze parameter iedere keer wordt gekozen bij iedere steekproef, waardoor er een inschatting kan worden gemaakt van de invloed van ψ_i op het huidige datapunt. Iedere schatting is gebaseerd op r steekproeven (voor EDPL kiezen Nazarov en Jarosz $r = 50$). Uiteindelijk wordt $P(X | \psi_i, G_t)$ geschat door het aantal matches, gegeven G'_t waarin $P(\psi_i) = 1$, te delen door r (zie 19a). Hetzelfde proces wordt herhaald voor $\neg\psi_i$, waarbij alle andere parameters ongewijzigd blijven, om zo de effecten van de bewuste parameter te kunnen isoleren.

Jarosz (2015) stelde voor dat de berekening van $R(\psi_i)$ af zou hangen van twee stappen: namelijk de waarschijnlijkheid van het huidige datapunt X schatten gegeven een specifieke parameterwaarde, i.e., $\hat{P}(X | \psi_i, G_t)$, en deze waarschijnlijkheid via Bayes' regel⁸ omzetten naar $P(\psi_i | \psi_i, G_t)$. Om delen door nul te voorkomen, wat voor kan komen als de kans op een parameterwaarde 0 is, wordt in de implementatie een zeer klein getal $\epsilon = 10^{-250}$ toegevoegd aan het aantal matches voor iedere ψ_i . Verdere toepassing van Bayes' regel en de wet van de totale kans⁹ levert dan (19c), precies wat we zoeken: de verwachte waarschijnlijkheid op een match met datapunt X gegeven grammatica G_t .

De eerder beschreven problemen bij NPL, het lifter- en handlangersscenario, vormen geen moeilijkheid bij de EDPL. De beloningswaarden worden specifiek voor iedere parameterwaarde berekend en zijn wél gevoelig voor de invloed van de parameterwaarde in kwestie naar het datapunt. Voor specifieke voorbeelden zie Nazarov en Jarosz (2017).

⁸ $P(A | B) = \frac{P(B|A)P(A)}{P(B|A)P(A)+P(B|\neg A)P(\neg A)}$

⁹ $P(B) = P(B | A)P(A) + P(B | \neg A)P(\neg A)$

Iedere run wordt voor iedere parameter tweemaal (ψ_i en $\neg\psi_i$) uitgevoerd. Dat komt neer op $2 \times 50 = 100$ iteraties per parameter in G_t .

3.3 Datasets

Om de frequentie van zware lettergrepen eenvoudig te kunnen manipuleren, gebruiken we een kunsttaal die louter bestaat uit combinaties van een beperkte verzameling aan lettergrepen. Nazarov en Jarosz (2017) gebruiken de verzameling {"ta", "tan", "taa"} om woorden uit te genereren. "ta" is altijd een lichte lettergreep, "tan" is licht of zwaar afhankelijk van of de taal kwantiteitssensitief in de rijm of nucleus is (P5, CVC) en "taa" is altijd zwaar. De dataset bestaat uit alle combinaties van minimaal drie en maximaal zes lettergrepen die te verkrijgen zijn uit bovengenoemde atomaire verzameling (alfabet) met een eenvoudig algoritme.¹⁰ De lettergrepen worden van elkaar gescheiden door middel van een koppelteken, e.g. "ta-ta-ta". Deze dataset met alle combinaties noemen we D_0 . Het alfabet noteren we als Σ_i . Dus $\Sigma_0 = \{\text{"ta"}, \text{"tan"}, \text{"taa"}\}$ en $D_0 = \{\text{"ta-ta-ta"}, \text{"ta-ta-tan"}, \text{"ta-ta-taa"}, \dots, \text{"ta-ta-ta-ta"}, \text{"ta-ta-ta-tan"}, \text{"ta-ta-ta-taa"}, \dots, \text{"taa-taa-taa-taa-taa-taa"}\}$. Dat komt neer op $3^3 + 3^4 + 3^5 + 3^6 = 1,080$ datapunten.

In Tabel 2 staat een overzicht van alle datasets, met bijbehorende atomaire elementen en distributies. Door het alfabet Σ_0 uit te breiden met {"taan", "tee", "teen"} verkrijgen we Σ_1 . Dit zijn allemaal zware lettergrepen. Deze nieuwe verzameling heeft nu een verhouding 3:1 van zware tot lichte lettergrepen (Z:L) voor talen waarbij CVC-lettergrepen als licht gelden (CVC=LIGHT). Bij de andere waarden van de CVC-parameter zien we een verhouding 6:1 ten faveure van de zware lettergrepen.

Een tweede uitbreiding van Σ_0 met {"tee", "too", "tuu"} resulteert in Σ_2 . Het verschil met Σ_2 is de evidentie dat een lettergreep eindigend op een consonant, i.e., een coda al dan niet als zwaar wordt beschouwd. Deze verzameling heeft verder dezelfde distributie van zware lettergrepen als Σ_1 .

D_0	ta	tan	taa				Z:L
	CV	CVC	CVV				
	L	L/Z	Z				3:1 of 1:3
D_1	ta	tan	taa	taan	tee	teen	
	CV	CVC	CVV	CVVC	CVV	CVVC	
	L	L/Z	Z	Z	Z	Z	3:1 of 6:1
D_2	ta	tan	taa	tee	too	tuu	
	CV	CVC	CVV	CVV	CVV	CVV	
	L	L/Z	Z	Z	Z	Z	3:1 of 6:1

TABEL 2: Datasets

D_1 en D_2 bevatten ieder $6^3 + 6^4 + 6^5 + 6^6 = 55,944$ datapunten. We kruisen iedere dataset D_i met iedere grammatica G_i (zie § 3.4) om voor ieder element een klemtoonpatroon te genereren.

In (17) zagen we al een voorbeeld van hoe een klemtoonpatroon gerepresenteerd wordt als string. Een klemtoonstring, bijvoorbeeld "20100", geeft de posities aan van de lettergrepen die een klemtoon krijgen. "2" representeert een hoofdklemtoon, "1" een nevenklemtoon en "0" geldt als onbeklemtoond.

¹⁰Voor n lettergrepen heb je $n - 1$ geneste for-loops nodig of een recursieve equivalent.

3.4 Grammatica's

In het experiment onderzoeken we vier grammatica's. Twee grammatica's zijn problematisch volgens Nazarov en Jarosz (2020), i.e., die convergeren niet met EDPL. Dit zijn $\dagger G_1$ en $\dagger G_3$.¹¹ Deze grammatica's verschillen van G_0 respectievelijk G_2 in de parameter CVC; de problematische talen hebben CVC = HEAVY. Daarnaast verschillen ($G_0, \dagger G_1$) met ($G_2, \dagger G_3$) in of de voeten van links of rechts worden opgebouwd (DIR) en of de voeten jambisch of trocheïsch zijn (FOOT) en of de klemtoon van links of rechts wordt gegeven (MAIN). Dit is overzichtelijk weergegeven in Tabel 3.

	XM	XMDIR	DIR	BOUNDED	FOOT	MAIN	QS	CVC	HEAVYHEAD	SECSTRESS	DEGENERATE
G_0	OFF	RIGHT	L-TO-R	ON	IAMB	LEFT	ON	LIGHT	OFF	ON	OFF
$\dagger G_1$	OFF	RIGHT	L-TO-R	ON	IAMB	LEFT	ON	HEAVY	OFF	ON	OFF
G_2	OFF	RIGHT	R-TO-L	ON	TROCHEE	RIGHT	ON	LIGHT	OFF	ON	OFF
$\dagger G_3$	OFF	RIGHT	R-TO-L	ON	TROCHEE	RIGHT	ON	HEAVY	OFF	ON	OFF

TABEL 3: Grammatica's

Laten we bij wijze van een voorbeeld datapunt "tan-taa-taa-tan" (voorkomend in alle datasets) beschouwen per grammatica.

(20)	a. G_0	b. $\dagger G_1$	c. G_2	d. $\dagger G_3$
	('tan)(taa)(taa)(tan)	(tan'taa)(taa)tan	(tan.taa)(taa'tan)	tan(taa)('taa.tan)
	"2000"	"0200"	"0002"	"0020"

Alle grammatica's leiden tot een opdeling in metrische voeten met een maximale grootte van twee lettergrepen, immers BOUNDED=ON. Dat de voeten van links naar rechts worden opgebouwd maakt in dit geval niet uit, omdat we te maken hebben met een symmetrisch datapunt met een even aantal lettergrepen.

Als we de grammatica's uit (20a) en (20b) vergelijken, zien we alleen verschil in CVC. De voeten zijn dus kwantiteitssensitief in de rijm respectievelijk de nucleus. De syllabe "tan" is van het type CVC en heeft dus een coda. Bij kwantiteitssensitiviteit in de rijm wordt dat dus een zware lettergreep en daarmee het hoofd van de voet. Omdat de hoofdklemtoon MAIN bepaald wordt van links naar rechts, wordt (tan.taa) de hoofdvoet en "tan" daarmee beklemtoond.

In (20b) geldt "tan" dus niet als zwaar en is "taa" de enige zware lettergreep uit de voet. Hiermee wordt "taa" het hoofd van de voet en krijgt het dus de hoofdklemtoon. Omdat DEGENERATE=OFF valt de tweede "tan" (L) buiten de versvoet.

Hetzelfde proces speelt zich af van rechts naar links voor (20c) en (20d). Daarmee krijgt "tan" voor (20c) de hoofdklemtoon en "taa" voor (20d).

Merk op dat er geen nevenklemtonen worden aangewezen omdat SECSTRESS=OFF. Verder maken de parameters BOUNDED en DIR in de specifieke voorbeelden uit 3.4 niet uit. Ga na dat dat anders zou zijn bij datapunten met een andere lengte of met een andere zwaarte van de lettergrepen.

¹¹De \dagger geeft aan dat het een problematische grammatica betreft.

4 Resultaten

Nazarov en Jarosz (2017) implementeerden EDPL in R¹². De input bestond uit alle elementen uit alle datasets D_i met het gegenereerde stresspatroon uit alle grammatica's G_i . Voor iedere leerder werden er 10 runs gedaan en iedere run bleef 100.000 iteraties doorgaan of stopte bij convergentie, waarvoor iedere 100 iteraties werd gecheckt. We nemen de definitie van een succesvolle convergentie over van Nazarov en Jarosz (2017), i.e., een run convergeert als de resulterende grammatica het klemtoonpatroon 99 van de 100 keer juist heeft. De resultaten zijn samengevat in Tabel 4.

	D_0	D_1	D_2
G_0	10/10	10/10	10/10
$\dagger G_1$	0/10	10/10	10/10
G_2	10/10	10/10	10/10
$\dagger G_3$	0/10	10/10	10/10

TABEL 4: Succesvolle convergenties

Het aantal iteraties per run dat een leerder nodig heeft om te convergeren is een factor in het bepalen van de kwaliteit van de leerder, i.e., een snellere leerder is beter. In Tabel 5 is een overzicht gegeven van het aantal iteraties per kruising. Vanwege de kleine hoeveelheid data, $n = 10$ voor iedere run, heb ik in plaats van het gemiddelde de mediaan gegeven.

		D_0	D_1	D_2
G_0	med.	600	1750	1800
	min-max	600-700	1500-2000	1600-2100
$\dagger G_1$	med.	N/A	600	600
	min-max	N/A	500/700	500/700
G_2	med.	650	1700	1800
	min-max	500-800	1200-2400	1200-2400
$\dagger G_3$	med.	N/A	600	600
	min-max	N/A	500-700	400-700

TABEL 5: Iteraties

Voor de volledige resultaten, zie Appendix B. Uit de data kunnen we opmaken dat de problematische talen $\dagger G_1$ en $\dagger G_3$ niet convergeren bij de basisdataset D_0 . Bij de datasets waarin zware lettergrepen vaker voorkomen, worden de problematische talen wel geleerd. Verder zien we dat grammatica's met CVC=HEAVY, i.e., de problematische talen, sneller geleerd worden dan de talen met CVC=LIGHT.

4.1 Dataset met lichte lettergrepen

Om te testen wat er gebeurt bij een lage frequentie van zware lettergrepen in een dataset met CVC=HEAVY is een extra kunsttaal nodig. Daarvoor kunnen we alfabet Σ_0 uit § 3.3 uitbreiden naar $\Sigma_3 = \Sigma_0 \cup \{"te", "to", "tu"\}$. In dat geval hebben we een distributie van 1:3 (Z:L) bij CVC=LIGHT en anders 1:6. Hiermee kunnen we taal D_3 genereren.

¹²RStudio 1.3.959

D_3	ta	te	to	tu	tan	taa	Z:L
	CV	CV	CV	CV	CVC	CVV	
	L	L	L	L	L/Z	Z	1:3 of 1:6

TABEL 6: Dataset met hoge frequentie lichte lettergrepen

Onze hypothese volgend, zou de verwachting zijn dat zowel de taal met CVC=LIGHT als CVC=HEAVY niet convergeert met EDPL. De voorlopige resultaten onderschrijven dat (zie ook Appendix B.1). Met G_0 convergeerde EDPL op D_3 0 uit 2 runs. Voor ${}^\dagger G_1$, G_2 en ${}^\dagger G_3$ was dat 0 uit 1 run. In vervolgonderzoek zou nader op een distributie waarin lichte lettergrepen frequent voorkomen in kunnen worden gegaan.

5 Discussie

5.1 Leerbaarheid

In de resultaten (Tabel 4) komt zeer duidelijk naar voren dat de problematische grammatica's niet convergeren met EDPL voor basisdataset D_0 . Daarmee wordt de hypothese, i.e., de frequentie van zware lettergrepen hebben invloed op de leerbaarheid, bevestigd. Op de datasets waar de zware lettergrepen frequenter voorkomen worden diezelfde problematische namelijk talen namelijk wél geleerd, namelijk

De talen die niet convergeerden in EDPL volgens Nazarov en Jarosz (2020) convergeren hiermee wel. EDPL is in staat om met succes alle klemtoonsystemen te leren. Hiermee is het punt dat Pearl (2011) maakt, namelijk dat domeinalgemene leerders niet in staat zijn om P&P-klemtoonsystemen te leren zonder een bias, ontkracht.

Er is nog niet duidelijk wat de invloed is van een lage frequentie zware lettergrepen c.q. een hoge frequentie lichte lettergrepen. Voorlopige resultaten (zie Tabel 6) indiceren dat dat een negatieve impact heeft op de leerbaarheid van de problematische grammatica's van Nazarov en Jarosz (2020). De vraag of er een nieuwe schare aan problematische talen ontstaat met deze dataset, en zo ja, welke parameters daar dan mee gemoeid zijn blijft open.

5.2 Snelheid

Problematische grammatica's worden sneller geleerd op de datasets met veel zware lettergrepen dan de niet-problematische grammatica's, zie Tabel 5. Zo worden $\dagger G_1$ en $\dagger G_3$ op D_1 met $\alpha = 1\%$ significant sneller geleerd dan G_0 en G_2 (t-toets, $p < 0,01$; $df = 38$). Hetzelfde effect treedt op tussen problematische en niet-problematische grammatica's op D_2 ($p < 0,01$; $df = 38$).

Daarnaast convergeren niet-problematische grammatica's eerder op datasets zonder hogere frequentie zware lettergrepen; G_0 en G_2 worden significant ($\alpha = 1\%$) sneller geleerd op D_0 dan op D_1 (t-toets, $p < 0,01$; $df = 38$). Dit gaat ook op voor de vergelijking tussen D_0 en D_2 (t-toets, $p < 0,01$; $df = 38$).

Hierbij wil ik de kanttekening plaatsen dat er weinig datapunten zijn; $n = 10$ per grammatica, en dus $N = 20$ voor de grammatica's die samen onder de categorie 'problematisch' danwel 'niet-problematisch' worden geschaard. Er schuilen achter een datapunt wel 100.000 iteraties als er niet wordt geconvergeerd (time-out waarde). Voor een beter beeld zouden er meer runs moeten worden uitgevoerd tegen een hogere time-out waarde.

6 Conclusie

Dit onderzoek begon met de vraag of de frequentie van zware lettergrepen van invloed was op de leerbaarheid van door Nazarov en Jarosz (2020) aangewezen, problematische talen in domeinalgemene leerders. Deze vraag is relevant in de context van de acquisitie van klemtoonsystemen. Pearl (2011) stelde dat leerders van dergelijke systemen alleen succesvol konden zijn als er, naast de P&P, een bias zou worden ingebouwd. Daaropvolgend lieten Nazarov en Jarosz (2017) zien dat met hun Expectation Driven Parameter Learner (EDPL) een veel beter resultaat te behalven viel dan die gepresenteerd door Pearl; maar EDPL convergeerde niet in alle gevallen. Een vervolgstudie liet zien om welke grammatica's het ging (Nazarov en Jarosz 2020). Het bleek dat er voorbeelden van natuurlijke talen waren die correspondeerden met de gevonden grammatica's, namelijk Frans en Polabisch. Om Pearls stelling dat onbevooroordeelde leerders niet zonder meer in staat zijn om klemtoonsystemen te leren te kunnen weerspreken, moest EDPL in staat zijn ook voor die talen te convergeren.

Omdat de grammatica's in kwestie, kwantiteitssensitief aan de nucleus, tot een bepaalde distributie van zware en lichte lettergrepen leidde, heb ik onderzocht of die eigenschap invloed had op de leerbaarheid. Daarvoor nam ik de dataset van Nazarov en Jarosz (2020), waarop EDPL níet convergeerde met de problematische grammatica's. Vervolgens breidde ik dat uit met twee andere datasets, waarbij de frequentie van zware lettergrepen hoger lag. Het bleek dat deze talen wél convergeerden op de nieuwe datasets; met het convergeren van deze laatste, problematische talen heb ik laten zien dat EDPL in staat is om alle klemtoonsystemen te leren. EDPL, een onbevooroordeelde, domeinalgemene leerder, vormt dus een tegenbeeld voor de conclusie die Pearl trekt.

Naast de effecten op de leerbaarheid, heb ik ook effecten van de lettergreepdistributies op de snelheid van de leerder gevonden. Zo worden de problematische grammatica's sneller geleerd op de datasets met veel zware lettergrepen dan de niet-problematische grammatica's. Er trad ook een effect op tussen niet-problematische grammatica's en datasets die geen hoge frequentie zware lettergrepen hadden; die grammatica's werden op die datasets namelijk sneller geleerd dan op de datasets met een hoge frequentie zware lettergrepen. Er is dus een verband tussen de parametrische eigenschappen van een grammatica enerzijds, en de distributie van zware lettergrepen in de dataset anderzijds als het gaat om de snelheid van de convergentie.

Dit kan een interessant aanknopingspunt voor vervolgonderzoek zijn; worden talen die kwantiteitssensitief in de nucleus zijn, of andere parametrische overeenkomsten hebben met de probleemtalen, sneller geleerd op datasets met een hogere frequentie zware lettergrepen? Hangt dat uitsluitend van die factoren af? Zijn er meer combinaties tussen bepaalde distributies en de snelheid waarmee een leerder convergeert en zo ja, welke implicaties hebben die op ons psychologische begrip van taalverwerving? Inherent aan het P&P-framework is de co-invloed van parameters op de resultaten, wat het lastig maakt om de invloed van één parameter te isoleren.

Ten slotte heb ik ook wat voorlopige resultaten voor een dataset met een lage frequentie aan zware lettergrepen, c.q. een hoge frequentie aan lichte lettergrepen. Deze resultaten indiceren dat geen enkele taal geleerd wordt met EDPL bij gebruik van die dataset. Wellicht bestaan er wel grammatica's die wel geleerd worden op die dataset. Het zou ook kunnen zijn dat de convergentiesnelheid positief beïnvloed

wordt voor bepaalde grammatica's. De invloed van een *lage* frequentie aan zware lettergrepen is een andere interessante invalshoek.

Al met al kan ik zeggen dat er nog veel ruimte is voor nader onderzoek over de invloed van de lettergreepdistributies en de parametrische verschillen tussen grammatica's op de leerbaarheid in EDPL.

Referenties

- Allen, W. Sydney. 1973. *Vox Latina. The pronunciation of Classical Latin*. 2de ed. Cambridge University Press.
- Apoussidou, Diana. 2007. „The learnability of metrical phonology.” Proefschrift, University of Amsterdam.
- Bailey, Todd M. 1995. „Nonmetrical constraints on stress”. Masterscriptie, University of Minnesota, Minneapolis.
- Bethin, Christina Y. 1998. *Slavic Prosody: Language Change and Phonological Theory*. Cambridge studies in Linguistics, 86. Cambridge: Cambridge University Press.
- Bush, R.R., en F. Mosteller. 1951. „A model for stimulus generalization and discrimination”. *Psychological Review* 58 (6): 413–423.
- Chomsky, N. 1986. *Knowledge of language: its nature, origin, and use*. Convergence Series. Praeger.
- Chomsky, Noam. 1981/1982. *Lectures on government and binding. Second revised edition*. Dordrecht: Foris.
- Chomsky, Noam. 1957. *Syntactic structures*. The Hague: Mouton.
- Clark, Robin. 1990. „Papers on learnability and natural selection”. In *Technical Reports in Formal and Computational Linguistics, No 1*. Université de Genève.
- Clark, Robin. 1992. „The selection of syntactic knowledge”. *Language Acquisition* 2:83–149.
- Dempster, A. P., N. M. Laird en D. B. Rubin. 1977. „Maximum Likelihood from Incomplete Data via the EM Algorithm”. *Journal of the Royal Statistical Society. Series B (Methodological)* 39 (1): 1–38.
- Demuth, Katherine. 1996. „The prosodic structure of early words”. In *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition*, red. door James Morgan en Katherine Demuth, 171–184. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Dresher, B. Elan, en Jonathan D. Kaye. 1990. „A computational learning model for metrical phonology”. *Cognition* 34 (2).
- Dresher, Elan. 1999. „Charting the learning path: Cues to parameter setting”. *Linguistic Inquiry* 30 (1): 27–67.
- Ewen, Colin J., en Harry van der Hulst. 2001. *The phonological structure of words: an introduction (Cambridge Textbooks in Linguistics)*. Cambridge Univ. Press.
- Fikkert, Paula. 1994. *On the acquisition of prosodic structure*. Dordrecht: Holland Institute of Generative Linguistics.
- Fodor, Janet D. 1998. „Unambiguous triggers”. *Linguistic inquiry* 29 (1): 1–36.
- Goedemans, Rob W., Jeffrey Heinz en Harry van der Hulst. 2015. „StressTyp2, versie 1”. <http://st2.ullet.net>.
- Hayes, Bruce. 1981. *A Metrical Theory of Stress Rules*. Indiana University Linguistics Club.
- Hayes, Bruce. 1989. „Compensatory Lengthening in Moraic Phonology”. *Linguistic Inquiry* 20 (2): 253–306. ISSN: 00243892, 15309150.

- Hayes, Bruce. 1995. *Metrical Stress Theory: Principles and Case Studies*. University of Chicago Press.
- Hueting, A. 1936. „Iets over de Spraakkunst van de Tobèloreesche Taal”. *Bijdragen tot de taal-, land- en volkenkunde* 94 (1): 295–407.
- van der Hulst, Harry, en Michael Moortgat. 1980. „Prosodische fonologie en de accentuatie van Nederlandse woorden, of: Leeft het Nederlands op grote voet?” *Verslag van de 150e vergadering van de vereniging van fonetische wetenschappen* 1–25.
- Jarosz, Gaja. 2015. *Expectation Driven Learning of Phonology*. Ms. University of Massachusetts Amherst.
- Levelt, Clara C., Niels O. Schiller en Willem J. Levelt. 2000. „The acquisition of syllable types”. *Language Acquisition* 8 (3): 237–64.
- Lightfoot, David. 1999. *The development of language: Acquisition, change, and evolution*. Oxford: Blackwell.
- Maan, G. 1953. „Proeve van een Bulische spraakkunst”. In *Verhandelingen van het Koninklijk Instituut voor Taal-, Land- en Volkenkunde*, 10. Leiden: Brill.
- McCarthy, John, en Alan Prince. 1993. „Prosodic Morphology I: Constraint Interaction and Satisfaction”. *Linguistics Department Faculty Publication Series* 14.
- Nazarov, Aleksei, en Gaia Jarosz. 2020. *Domain General Learning of Parametric Stress*. *Artikel in voorbereiding*. Ms., Utrecht University & University of Massachusetts Amherst.
- Nazarov, Aleksei, en Gaja Jarosz. 2017. „Learning Parametric Stress without Domain-Specific Mechanisms”. *Proceedings of the Annual Meetings on Phonology* 4.
- Nespor, Marina, en Irene Vogel. 1986. *Prosodic Phonology*. Studies in generative grammar. Dordrecht: Foris.
- van Oostendorp, Marc. 2005. „Mora Theory”. <http://www.vanoostendorp.nl/pdf/051122.pdf>.
- Pearl, Lisa S. 2009. „Acquiring complex linguistic systems from natural language data: What selective learning biases can do”. Ms. *University of California, Irvine*.
- Pearl, Lisa S. 2008. „Putting the Emphasis on Unambiguous: The Feasibility of Data Filtering for Learning English Metrical Phonology”, red. door Harvey Chan, Heather Jacob en Enkeleida Kapia, 390–401. Somerville, MA: Cascadilla Press.
- Pearl, Lisa S. 2011. „When Unbiased Probabilistic Learning Is Not Enough: Acquiring a Parametric System of Metrical Phonology”. *Language Acquisition* 18 (2): 87–120.
- Pearl, Lisa S., en Amy Weinberg. 2007. „Input filtering in syntactic acquisition: Answers from language change modeling”. *Language Learning and Development* 3 (1): 43–72.
- Ryan, Kevin M. 2016. „Phonological weight”. *Language and Linguistics Compass* 10 (12): 720–733.
- Selkirk, Lisa. 1978. „The French Foot: On the Status of Mute e”. *Studies in French Linguistics* 1 (2): 141–150.
- Stone, Gerald. 1972. „The Smallest Slavonic Nation: The Sorbs Of Lusatia”. London: Athlone Press of the University of London.
- Yang, C.D. 2002. *Knowledge and Learning in Natural Language*. Oxford University Press.

A Parameters

Dresher en Kaye (1990) beschrijven elf parameters van klemtoonssystemen in P&P. Hieronder staat het overzicht, met achter iedere formulering de afkorting zoals in Nazarov en Jarosz (2017).

- (21) P1: The word-tree is strong on [Left/Right] (MAIN=LEFT/RIGHT)
- P2: Feet are [Binary/Unbounded] (BOUNDED=ON/OFF)
- P3: Feet are built from the [Left/Right] (DIR=L-TO-R/R-TO-L)
- P4: Feet are strong on the [Left/Right] (FOOT=TROCHEE/IAMB)
- P5: Feet are quantity sensitive (QS) [Yes/No] (QS=ON/OFF)
- P6: Feet are QS to the [Rime/Nucleus] (CVC=LIGHT/HEAVY)
- P7: A strong branch of a foot must itself branch [No/Yes] (HEAVYHEAD=ON/OFF)
- P8A: There is an extrametrical syllable [No/Yes] (XM=ON/OFF)
- P8: It is extrametrical on the [Left/Right] (XMDIR=LEFT/RIGHT)
- P9: Feet consisting of a single light syllable are removed [No/Yes] (DEGENERATE=ON/OFF)
- P10: Feet are noniterative [No/Yes] (SECSTRESS=ON/OFF)

Dresher en Kaye spreken van *cues* om parameters te herkennen in talen. Zo'n *cue*, aanwijzing, is dan een eigenschap die duidt op het voorkomen van een parameter. Per parameter zal ik een korte uitleg geven, Dresher en Kaye parafraserend. Voor een overzicht van de cues, die minder relevant zijn voor dit onderzoek, verwijst ik naar Dresher en Kaye (1990, pp. 190–194)

P1: hoofdbeklemtoning. Deze parameter geeft aan of het toewijzen van de hoofdklemtoon plaatsvindt vanaf de linker- of de rechterkant van een woord. Bij een ongebonden taal wordt de hoofdklemtoon zo direct bepaald door deze parameter, i.e., de eerste of de laatste lettergreep. De hoofdklemtoon is niet per se gebonden aan een uiterste lettergreep, omdat P1 ook interacteert met andere parameters, e.g. P8A/P8: een extrametrische lettergreep kan niet de hoofdklemtoon krijgen.

P2: gebonden en ongebonden voeten. Binaire (gebonden) voeten zijn volgens Dresher en Kaye van de vorm (LL), (H), (HL) of (LH). Ongebonden voeten hebben geen maximale grootte. Bij talen met gebonden voeten speelt de positie de grootste factor in het bepalen van de klemtoon, terwijl dat bij ongebonden voeten meer afhangt van het lettergreepgewicht.

P3–P4: constructie van voeten. Deze parameters hebben betrekking op gebonden voeten (P2). Voeten worden geconstrueerd van links naar rechts, of van rechts naar links, e.g. (ta.ta)(ta.ta)ta v.s. ta(ta.ta)(ta.ta). Daarnaast kan het accent in een voet links of rechts liggen, ('ta.ta) vs. (ta.'ta).

P5–P6: kwantiteitssensitiviteit. Een taal is kwantiteitssensitief als de voeten afhankelijk zijn van de interne structuur van de lettergreep, i.e. het hoofd van een voet is altijd een zware lettergreep; bij kwantiteitsinsensitiviteit kan het hoofd van een voet zowel licht als zwaar zijn. Daarnaast kan een taal kwantiteitssensitief in de *rijm* of in de *nucleus* zijn (P6). Daarvoor gelden gesloten lettergrepen, i.e., van de vorm CVC, respectievelijk als licht of zwaar.

P7: zware lettergrepen als hoofd van voeten. Deze parameter geeft (bij HEAVYHEAD=ON) aan dat alleen voeten overblijven waarbij het hoofd een zware lettergreep is. Dat betekent dat lichte lettergrepen die geen onderdeel uitmaken van een voet, geen aparte

voet kunnen vormen, e.g. ta.ta.taa.ta.taa \rightarrow (ta.ta)(taa.ta)(taa) \rightarrow ta.ta(taa.ta)(taa)
 Bij kwantiteitsinsensitieve talen resulteert dit in het verwijderen van alle voeten. In dat geval wordt de klemtoon bepaald door P1.

P8A–P8: extrametricaliteit. In een extrametrical taal wordt de eerste of laatste lettergreep niet meegerekend in een metrische voet. Zo'n extrametrical lettergreep kan dus nooit beklemtoond worden. De parameter P8 geeft aan of dat de eerste (LEFT) of laatste (RIGHT) lettergreep genegeerd wordt. Merk op dat als XM=OFF de waarde van P8 geen invloed meer heeft.

P9: geen (L). Voeten met alleen een lichte lettergreep kunnen niet voorkomen met deze parameter.

P10: geen nevenklemtonen. Talen zonder secundaire stress krijgen één klemtoon, de hoofdklemtoon, toegewezen. Alleen de sterkste voet krijgt de klemtoon en de rest blijft onbeklemtoond.

B Volledige resultaten

Op de volgende pagina's staan de resultaten voor D_0 , D_1 en D_2 . Alle waarden zijn afgerond op drie decimalen. De eerste kolom bestaat uit het nummer van de run: 1 t/m 10. Dan volgen er elf kolommen die de uiteindelijke waarde van de parameters aangeven. Daarbij is de oorspronkelijke naamgeving van de parameterwaarden uit EDPL (Nazarov en Jarosz 2017) aangehouden. Een voorbeeld: een kans van 0,179 op "XM" betekent: $P(\text{XM:LEFT}) = 0,179$ en consequent $P(\text{XM:RIGHT}) = 1 - 0,179 = 0,821$. Zie (21) voor een overzicht.

De voorlaatste kolom geeft het aantal iteraties tot convergentie aan of, als er geen convergentie heeft plaatsgevonden, de time-outwaarde (1e+05). De laatste kolom, *Last_mismatch*, is een benadering van wanneer de leerder voor het laatst met een bepaalde waarschijnlijkheid een verkeerde vorm produceerde; het is een schatting die bedoeld is om te kijken of er vaak genoeg voor convergentie wordt gecheckt.

#	XM	XML	L.to.R	Bounded	Trochee	Leftmost	QS	QS.CVC	OblBr	NonIt	Defooting	Iterations	Last_mismatch
$D_0 \times G_0$													
1	0.179	4.572e-05	0.997	1.000	1.713e-05	1.000	1.000	1.000	0.001	1.000	0.680	700	612
2	0.001	0.096	0.994	1.000	3.111e-06	1.000	1.000	1.000	0.000	1.000	0.653	700	563
3	0.005	0.110	0.986	0.999	0.001	1.000	1.000	1.000	0.003	1.000	0.751	500	482
4	7.106e-05	0.185	0.994	0.999	6.024e-05	1.000	1.000	1.000	0.002	1.000	0.717	600	491
5	0.019	0.022	0.997	1.000	1.905e-05	1.000	1.000	1.000	0.000	1.000	0.769	700	514
6	0.145	0.000	0.995	1.000	5.729e-05	1.000	1.000	1.000	0.001	1.000	0.707	800	651
7	0.001	0.120	0.988	0.999	0.000	1.000	1.000	1.000	0.005	1.000	0.637	700	645
8	0.015	0.028	0.996	1.000	4.677e-05	1.000	1.000	1.000	0.002	1.000	0.643	700	619
9	6.230e-05	0.258	0.991	0.999	0.000	1.000	1.000	1.000	0.003	1.000	0.630	500	479
10	0.013	0.038	0.989	0.998	0.001	1.000	1.000	1.000	0.005	1.000	0.757	600	581
$D_0 \times \dagger G_1$													
1	0.922	0.921	0.557	0.407	0.893	1.000	2.470e-323	0.428	1.000	0.426	0.316	1e+05	1e+05
2	0.771	0.748	0.557	0.484	0.531	1.000	1.534e-278	0.498	1.000	0.569	0.612	1e+05	1e+05
3	0.806	0.778	0.513	0.287	0.406	1.000	2.326e-144	0.487	1.000	0.446	0.378	1e+05	1e+05
4	0.876	0.828	0.687	0.843	0.678	1.000	1.389e-93	0.542	1.000	0.482	0.533	1e+05	1e+05
5	0.806	0.724	0.503	0.503	0.604	1.000	6.589e-13	0.245	1.000	0.522	0.717	1e+05	1e+05
6	0.792	0.707	0.316	0.386	0.560	1.000	2.470e-323	0.579	1.000	0.464	0.390	1e+05	1e+05
7	0.806	0.786	0.302	0.475	0.588	1.000	6.582e-130	0.353	1.000	0.623	0.430	1e+05	1e+05
8	0.836	0.878	0.585	0.512	0.361	1.000	4.741e-285	0.484	1.000	0.621	0.291	1e+05	1e+05
9	0.845	0.841	0.601	0.547	0.447	1.000	2.470e-323	0.737	1.000	0.247	0.653	1e+05	1e+05
10	0.745	0.944	0.508	0.462	0.293	1.000	1.546e-185	0.616	1.000	0.383	0.342	1e+05	1e+05
$D_0 \times G_2$													
1	0.012	0.929	0.006	0.999	0.999	2.428e-23	1.000	1.000	0.004	1.000	0.752	500	475
2	0.194	1.000	0.010	0.999	1.000	3.793e-23	1.000	1.000	0.003	1.000	0.652	500	483
3	0.313	1.000	0.003	1.000	1.000	4.327e-32	1.000	1.000	0.001	1.000	0.790	700	537
4	0.111	0.999	0.011	0.999	0.999	1.467e-31	1.000	1.000	0.002	1.000	0.827	700	685
5	0.110	0.999	0.005	1.000	1.000	4.691e-32	1.000	1.000	0.001	1.000	0.778	700	688
6	0.119	0.998	0.003	0.997	1.000	1.252e-31	1.000	1.000	0.004	1.000	0.560	700	659
7	0.003	0.860	0.013	0.997	0.999	3.921e-23	1.000	1.000	0.008	1.000	0.660	500	491
8	0.388	1.000	0.016	0.998	0.999	1.982e-27	1.000	1.000	0.007	1.000	0.775	600	539
9	0.024	0.957	0.021	0.996	0.998	5.662e-23	1.000	1.000	0.009	1.000	0.751	500	496
10	0.138	0.999	0.003	0.999	1.000	3.069e-23	1.000	1.000	0.003	1.000	0.685	500	446
$D_0 \times \dagger G_3$													
1	0.628	0.308	1.588e-06	0.496	2.302e-127	4.053e-281	6.713e-111	0.321	0.629	1.000	0.487	1e+05	1e+05
2	0.722	0.244	0.545	0.624	0.418	2.470e-323	2.470e-323	0.537	1.000	0.452	0.521	1e+05	1e+05
3	0.837	0.111	0.472	0.751	2.686e-19	2.470e-323	0.451	0.556	2.662e-05	1.000	1.024e-145	1e+05	1e+05
4	0.795	0.372	0.335	0.659	0.505	2.470e-323	2.470e-323	0.554	1.000	0.596	0.467	1e+05	1e+05
5	0.591	0.094	0.523	0.637	0.576	2.470e-323	2.964e-323	0.513	1.000	0.446	0.479	1e+05	1e+05
6	0.561	0.060	0.377	0.478	0.445	2.470e-323	2.470e-323	0.485	1.000	0.502	0.314	1e+05	1e+05
7	0.850	0.197	0.422	0.372	0.711	2.470e-323	2.470e-323	0.739	1.000	0.546	0.502	1e+05	1e+05
8	0.871	0.271	1.463e-29	0.480	2.964e-323	2.470e-323	2.304e-11	0.336	0.743	1.000	0.649	1e+05	1e+05
9	0.921	0.396	0.447	0.530	0.419	2.470e-323	7.452e-15	0.421	1.000	0.558	0.389	1e+05	1e+05
10	0.757	0.169	0.595	0.577	0.482	2.470e-323	2.470e-323	0.452	1.000	0.452	0.418	1e+05	1e+05

TABEL 7: Resultaten voor D_0

#	XM	XML	L.to.R	Bounded	Trochee	Leftmost	QS	QS.CVC	OblBr	Nonlt	Defooting	Iterations	Last_mismatch
$D_1 \times G_0$													
1	0.002	0.019	0.995	0.997	7.223e-05	1	1	1	0.009	1	0.751	1800	NA
2	0.199	2.144e-13	0.987	0.995	0.001	1	1	1	0.014	1	0.859	1700	NA
3	2.844e-16	0.252	0.982	0.992	0.000	1	1	1	0.016	1	0.821	1700	NA
4	0.021	0.002	0.983	0.991	0.002	1	1	1	0.025	1	0.852	1500	NA
5	3.292e-05	0.058	0.994	0.998	6.815e-06	1	1	1	0.005	1	0.682	1800	NA
6	0.103	3.549e-08	0.994	0.998	1.247e-05	1	1	1	0.008	1	0.553	1800	NA
7	0.103	1.948e-08	0.989	0.994	0.001	1	1	1	0.012	1	0.883	1900	NA
8	0.096	1.348e-06	0.995	0.994	0.001	1.000	1.000	1.000	0.014	1.000	0.831	1700	1684
9	1.338e-20	0.336	0.992	0.995	0.001	1.000	1.000	1.000	0.014	1.000	0.823	1700	1690
10	0.267	5.606e-20	0.996	0.999	4.585e-05	1.000	1.000	1.000	0.004	1.000	0.815	2000	1518
$D_1 \times \dagger G_1$													
1	0.010	0.084	0.996	0.999	0.000	1.000	1.000	0.000	0.003	1.000	0.680	500	500
2	0.001	0.126	0.997	1.000	2.072e-05	1.000	1.000	1.619e-05	0.001	1.000	0.680	700	522
3	0.003	0.101	0.992	0.999	0.000	1.000	1.000	9.764e-05	0.002	1.000	0.744	600	465
4	0.244	4.624e-06	0.992	1.000	0.000	1.000	1.000	1.831e-05	0.001	1.000	0.784	700	629
5	0.007	0.071	0.991	0.998	0.000	1.000	1.000	4.890e-05	0.005	1.000	0.770	600	598
6	0.073	0.006	0.992	0.999	0.000	1.000	1.000	8.077e-05	0.003	1.000	0.708	600	522
7	0.000	0.132	0.998	1.000	4.484e-05	1.000	1.000	3.004e-05	0.001	1.000	0.732	700	624
8	0.010	0.055	0.975	0.998	0.000	1.000	1.000	5.644e-05	0.004	1.000	0.661	600	583
9	0.006	0.079	0.997	0.999	0.000	1.000	1.000	6.666e-05	0.002	1.000	0.768	600	543
10	0.004	0.066	0.998	0.999	2.805e-06	1.000	1.000	8.923e-07	0.001	1.000	0.594	700	690
$D_1 \times G_2$													
1	0.116	1.000	0.007	0.998	1.000	1.232e-96	1.000	1.000	0.006	1.000	0.728	2100	2028
2	0.113	1.000	0.007	0.997	1.000	1.519e-87	1.000	1.000	0.009	1.000	0.772	1900	1722
3	4.942e-07	0.891	0.004	0.998	1.000	3.588e-69	1.000	1.000	0.010	1.000	0.812	1500	1411
4	2.125e-05	0.919	0.007	0.997	1.000	2.363e-64	1.000	1.000	0.008	1.000	0.785	1400	1386
5	0.231	1.000	0.005	0.997	1.000	6.385e-115	1.000	1.000	0.009	1.000	0.803	2500	2152
6	0.380	1.000	0.010	0.994	1.000	1.458e-55	1.000	1.000	0.016	1.000	0.700	1200	1186
7	9.690e-06	0.952	0.005	0.999	1.000	2.442e-101	1.000	1.000	0.003	1.000	0.933	2200	2174
8	0.165	1.000	0.011	0.992	1.000	6.196e-83	1.000	1.000	0.016	1.000	0.759	1800	1393
9	0.110	1.000	0.008	0.997	0.999	7.819e-74	1.000	1.000	0.008	1.000	0.854	1600	1472
10	2.364e-16	0.696	0.016	0.995	1.000	1.027e-64	1.000	1.000	0.013	1.000	0.708	1400	1203
$D_1 \times \dagger G_3$													
1	0.003	0.907	0.014	0.998	1.000	3.673e-28	1.000	5.314e-05	0.004	1.000	0.704	600	593
2	0.018	0.965	0.008	0.999	1.000	3.647e-28	1.000	0.000	0.002	1.000	0.641	600	572
3	0.010	0.953	0.003	1.000	1.000	1.153e-32	1.000	2.634e-05	0.001	1.000	0.657	700	599
4	0.172	1.000	0.008	0.999	1.000	4.098e-28	1.000	3.168e-05	0.002	1.000	0.824	600	595
5	0.189	1.000	0.006	0.999	1.000	4.164e-28	1.000	0.000	0.003	1.000	0.738	600	562
6	0.000	0.884	0.002	1.000	1.000	8.355e-33	1.000	2.006e-05	0.000	1.000	0.715	700	668
7	0.014	0.948	0.004	0.999	1.000	3.896e-28	1.000	4.204e-05	0.001	1.000	0.745	600	558
8	0.105	0.999	0.003	1.000	1.000	8.175e-33	1.000	7.162e-06	0.001	1.000	0.808	700	671
9	0.121	0.997	0.010	0.996	0.999	1.471e-23	1.000	0.000	0.014	1.000	0.656	500	485
10	7.052e-05	0.787	0.007	0.998	1.000	3.473e-28	1.000	5.901e-05	0.004	1.000	0.592	600	547

TABEL 8: Resultaten voor D_1

#	XM	XML	L.to.R	Bounded	Trochee	Leftmost	QS	QS.CVC	OblBr	Nonlt	Defooting	Iterations	Last_mismatch
$D_2 \times G_0$													
1	1.210e-13	0.158	0.995	0.998	8.013e-05	1.000	1.000	1.000	0.007	1.000	0.838	2100	1690
2	0.003	0.013	0.993	0.998	0.000	1.000	1.000	1.000	0.008	1.000	0.841	1900	1786
3	1.190e-12	0.168	0.985	0.995	0.001	1.000	1.000	1.000	0.011	1.000	0.924	1900	1757
4	0.001	0.025	0.984	0.992	0.000	1.000	1.000	1.000	0.019	1.000	0.790	1800	1595
5	0.115	3.671e-09	0.997	0.999	1.598e-05	1.000	1.000	1.000	0.002	1.000	0.784	1800	1717
6	9.990e-07	0.082	0.989	0.995	0.000	1.000	1.000	1.000	0.011	1.000	0.836	1800	1753
7	1.300e-14	0.224	0.994	0.997	0.001	1.000	1.000	1.000	0.009	1.000	0.851	1800	1735
8	0.047	4.657e-05	0.990	0.994	0.001	1.000	1.000	1.000	0.013	1.000	0.908	2000	1999
9	2.882e-13	0.222	0.993	0.997	0.000	1.000	1.000	1.000	0.008	1.000	0.790	1600	1513
10	0.060	1.163e-05	0.989	0.996	8.645e-05	1.000	1.000	1.000	0.010	1.000	0.733	1800	1643
$D_2 \times \dagger G_1$													
1	0.137	0.000	0.996	0.999	6.651e-06	1.000	1.000	1.207e-05	0.001	1.000	0.626	700	570
2	0.005	0.074	0.995	0.999	1.185e-05	1.000	1.000	0.000	0.002	1.000	0.470	600	574
3	3.806e-05	0.284	0.992	0.999	0.001	1.000	1.000	0.000	0.003	1.000	0.755	500	490
4	0.002	0.078	0.998	0.998	0.001	1.000	1.000	8.339e-05	0.004	1.000	0.851	700	666
5	0.001	0.133	0.997	0.999	9.046e-05	1.000	1.000	0.000	0.002	1.000	0.689	600	542
6	0.000	0.141	0.995	0.998	0.000	1.000	1.000	0.000	0.004	1.000	0.659	600	566
7	0.110	0.002	0.991	0.999	0.000	1.000	1.000	0.000	0.003	1.000	0.745	600	588
8	6.468e-07	0.347	0.990	0.997	0.000	1.000	1.000	6.644e-05	0.007	1.000	0.695	600	562
9	0.055	0.017	0.990	0.997	0.001	1.000	1.000	0.001	0.007	1.000	0.722	500	475
10	4.291e-06	0.231	0.995	1.000	2.000e-05	1.000	1.000	2.983e-05	0.001	1.000	0.686	700	650
$D_2 \times G_2$													
1	0.015	0.995	0.007	0.997	1.000	1.073e-64	1.000	1.000	0.010	1.000	0.728	1400	1337
2	7.255e-14	0.803	0.008	0.997	1.000	5.585e-83	1.000	1.000	0.007	1.000	0.745	1800	1635
3	2.300e-08	0.926	0.005	0.998	1.000	1.469e-110	1.000	1.000	0.004	1.000	0.924	2400	2246
4	0.100	1.000	0.006	0.997	1.000	5.353e-83	1.000	1.000	0.007	1.000	0.814	1800	1764
5	2.862e-17	0.773	0.009	0.996	0.999	4.218e-92	1.000	1.000	0.011	1.000	0.918	2000	1855
6	1.771e-15	0.774	0.011	0.997	1.000	7.338e-83	1.000	1.000	0.007	1.000	0.829	1800	1755
7	1.261e-10	0.845	0.009	0.996	1.000	6.314e-83	1.000	1.000	0.011	1.000	0.757	1800	1775
8	1.441e-10	0.870	0.010	0.996	0.999	1.303e-96	1.000	1.000	0.009	1.000	0.904	2100	1941
9	0.324	1.000	0.006	0.998	1.000	4.135e-69	1.000	1.000	0.008	1.000	0.588	1500	1464
10	0.115	1.000	0.015	0.991	1.000	1.252e-55	1.000	1.000	0.018	1.000	0.702	1200	1029
$D_2 \times \dagger G_3$													
1	0.081	0.994	0.010	0.997	1.000	5.742e-28	1.000	0.000	0.006	1.000	0.668	600	587
2	0.050	0.985	0.008	0.997	0.999	4.758e-28	1.000	0.001	0.007	1.000	0.789	600	579
3	0.000	0.841	0.007	0.999	1.000	3.528e-28	1.000	6.541e-05	0.002	1.000	0.590	600	597
4	2.972e-05	0.717	0.016	0.998	1.000	1.549e-23	1.000	0.001	0.005	1.000	0.619	500	404
5	0.000	0.836	0.004	0.999	1.000	3.564e-28	1.000	5.406e-05	0.003	1.000	0.602	600	547
6	0.001	0.906	0.005	0.999	1.000	1.113e-32	1.000	0.000	0.002	1.000	0.636	700	700
7	0.132	0.999	0.007	0.999	1.000	2.035e-32	1.000	5.309e-05	0.003	1.000	0.687	700	696
8	0.074	0.979	0.034	0.989	0.997	5.114e-19	1.000	0.003	0.025	1.000	0.722	400	398
9	0.072	0.994	0.015	0.998	0.999	5.251e-28	1.000	7.186e-05	0.004	1.000	0.851	600	550
10	0.008	0.940	0.008	0.999	1.000	2.488e-28	1.000	5.703e-05	0.002	1.000	0.656	600	467

TABEL 9: Resultaten voor D_2

B.1 Voorlopige resultaten D_3

#	XM	XM.L	L.to.R	Bounded	Trochee	Leftmost	QS	Q.S.CVC	ObI.Br	NonIt	Defooting	Iterations	Last_mismatch
$D_3 \times G_0$													
1	0.792	0.799	0.421	0.501	0.587	1.000	1.976e-93	0.632	1.000	0.560	0.320	1e+05	1e+05
2	0.491	0.932	0.515	0.513	0.560	1.000	2.964e-323	0.460	1.000	0.689	0.302	1e+05	1e+05
$D_3 \times \uparrow G_1$													
1	0.857	0.911	0.429	0.483	0.491	1.000	3.237e-24	0.445	1.000	0.287	0.302	1e+05	1e+05
$D_3 \times G_2$													
1	0.921	0.118	0.600	0.427	0.510	2.470e-323	1.232e-200	0.578	1.000	0.474	0.433	1e+05	1e+05
$D_3 \times \uparrow G_3$													
1	0.908	0.005	0.504	0.374	0.407	2.470e-323	1.763e-32	0.509	1.000	0.595	0.368	1e+05	1e+05

TABEL 10: Voorlopige resultaten voor D_3