Modelovanie efektu fonologického susedstva v akvizícii jazyka pomocou jednoduchej rekurentnej siete

Martin Takáč

Centrum kognitívnych vied, KAI FMFI Univerzity Komenského
Mlynská dolina, 842 48 Bratislava
takac@ii.fmph.uniba.sk

Abstrakt

Štúdie detského lexikálneho vývinu ukazujú, že deti si ako prvé osvojujú slová znejúce podobne ako veľa iných slov z jazyka, ktorému sú vystavené (pochádzajú z tzv. hustých fonologických susedstiev). Deti s normálnym rečovým vývinom si neskôr osvoja aj slová s málo fonologickými susedmi, zatiaľ čo deti s vývinovými problémami nie. V tomto článku prezentujeme výpočtový model osvojovania a produkcie slov, ktorý reprodukuje efekt fonologického susedstva, a pokúšame sa ho vysvetliť na základe analýzy vnútornej dynamiky rekurentnej neurónovej siete.

1 Úvod

Schopnosť naučiť sa správne používať slová je pre kognitívny vývin kľúčová, avšak nie všetky deti si osvojujú jazyk rovnako rýchlo. V anglicky hovoriacich krajinách ovláda väčšina dvojročných detí v priemere asi 300 slov [10], ale približne 10-20% detí v tomto veku ovláda menej ako 50 slov [6]. Pri skúmaní štruktúry osvojeného lexikónu detí s oneskoreným rečovým vývinom (ďalej ORV) sa zistilo, že tieto deti si dokážu osvojiť slová znejúce podobne ako veľa iných slov z jazyka, ktorému sú vystavené [8]. Takéto slová si ako prvé osvojujú aj deti s normálnym rečovým vývinom (ďalej NRV) [12], avšak tieto deti si neskôr osvoja aj iné slová, zatiaľ čo ORV majú s nimi problém. Pre skúmanie tohto javu sa zaviedla *miera fonologického susedstva* (angl. neighbourhood density, ďalej ND) – počet slov v jazyku, ktoré sa od daného slova odlišujú v najviac jednej foném (pridaním, vynechaním, alebo zámenou).

Prečo si deti osvojujú niektoré slová skôr ako iné? Aký mechanizmus spôsobuje, že slová s väčším ND sú ľahšie osvojiteľné? V tomto článku referujeme o pokuse zreprodukovať empirické výsledky pomocou výpočtového modelu osvojovania a produkcie slov a ponúkneme vysvetlenie tohto javu na základe analýzy dynamiky modelu. V nasledujúcich častiach predstavíme architektúru a trénovanie modelu a výsledky simulácií. V časti 5 analyzujeme vnútornú dynamiku modelu. V záverečnej diskusii sa pokúsime extrapolovať získaný vhľad na reálnu detskú populáciu.

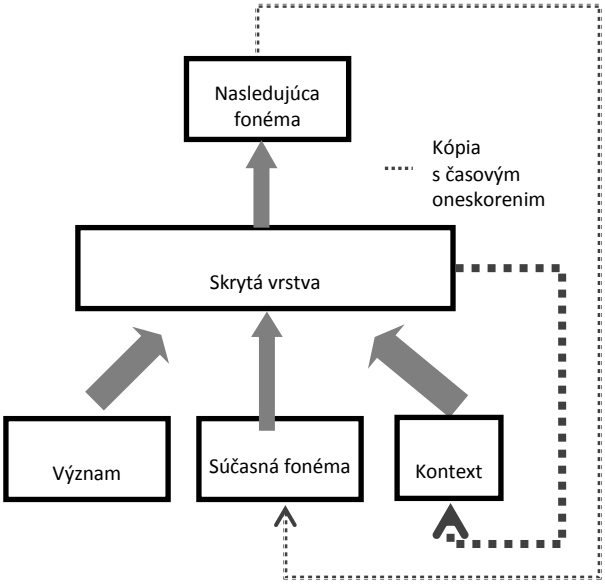
2 Architektúra modelu

Architektúra modelu (Obr. 1) pozostáva z jednoduchej rekurentnej siete (SRN) [3], ktorej vstup sa delí na reprezentáciu významu (*Význam*), poslednej vyslovenej fonémy (*Súčasná foném*a) a rekurentného kontextu (*Kontext*). Na výstupe sieť predikuje nasledujúcu fonému (*Nasledujúca foném*a). Fonémy sú reprezentované lokalisticky, vrstvy Súčasná aj Nasledujúca foném majú po 48 neurónov (47 pre všetky anglické fonémy a jeden neurón reprezentujúci hranicu medzi slovami – word boundary WB). Významy sú taktiež kódované lokalisticky – vrstva Význam má 268 neurónov pre 268 významov slov.

Počas trénovania sú sieti predkladané slová ako sekvencie jednotlivých foném a sieť sa učí predikovať pre daný význam slova a súčasnú fonému nasledujúcu fonému (a tiež na konci slova predikovať WB). Rekurentný kontext, ktorý je kópiou konfigurácie aktivít na skrytej vrstve z predchádzajúceho kroku, slúži na reprezentovanie histórie predchádzajúcich vstupov. Natrénovaná sieť dokáže naučené významy „vysloviť“ (teda vyprodukovať celú sekvenciu foném vhodného slova) tak, že na začiatku má na vstupe daný význam a WB a po každej predikcii sa predikovaná nasledujúca foném privedie naspäť na vstup, až kým model nepredikuje WB (alebo kým dĺžka generovanej postupnosti nepresiahne preddefinovanú rozumnú hranicu).

Kapacita SRN osvojovať si sekvencie súvisí s veľkosťou jej skrytej vrstvy [7]. Keďže jedna z hypotéz, prečo si ORV osvojujú slová pomalšie a v menšej miere, je, že môžu mať menšiu kapacitu fonologickej pracovnej pamäti [10], experimentovali sme s rôznymi verziami modelu, ktoré sa líšili veľkosťou skrytej vrstvy (5, 10, 15, 20 neurónov). Neuróny skrytej vrstvy mali sigmoidálnu aktivačnú funkciu, výstupne neuróny boli lineárne a skombinované softmax funkciou, takže súčet ich aktivít

bol 1 a mohli byť interpretované ako rozdelenie pravdepodobnosti možných nasledujúcich foném.



Obr. 1. Architektúra modelu. Hrubé šípky reprezentujú plne prepojené vrstvy (každý neurón s každým), bodkované šípky znamenajú kopírovanie s časovým oneskorením o jeden krok.

3 Trénovanie

Pri trénovaní nášho modelu vychádzame z predpokladu, že dieťa je vystavené veľkému množstvu slov: aj keď rozumie významu iba niektorých, aj z ostatných sa učí o fonologických zákonitostiach jazyka, prechodových pravdepodobnostiach medzi fonémami, atď. Na generovanie trénovacej množiny sme preto použili referenčnú databázu 2588 jednoslabičných anglických slov [2] získaných výberom z CELEX korpusu (17,9 miliónov slov) [1] vynechaním homofónov, homografov a skratiek.[1] Aby sme sa priblížili reálnemu jazykovému vstupu, trénovacia množina bola generovaná stochasticky tak, že pravdepodobnosť výskytu slova z databázy v trénovacej množine bola priamo úmerná jeho frekvencii výskytu (WF) v CELEX korpuse (presnejšie úmerná log(WF+1), aby sme dostali do trénovacej množiny aj menej frekventované slová). Trénovacia množina pozostávala z 2000 slov. Na určenie slov, ktorých významu by mohlo dieťa rozumieť, sme použili CDI inventár anglických slov [4]. Inventár obsahuje 672 slov a podľa štúdie [4] 50% detí vo veku 30 mesiacov ovláda z nich aspoň 600 slov. Z 2588 slov našej referenčnej databázy sa nachádza v CDI inventári 268, týmto sme v trénovacej množine priradili významy. Konkrétne, pokiaľ sa slovo z trénovacej množiny nachádza v CDI, počas trénovania na sekvencii foném tohto slova je na vrstve význam aktivovaný príslušný neurón. Počas trénovania na sekvenciách slov mimo CDI nie je na vrstve význam aktívny žiaden neurón. Napríklad pre slovo „*dog*“ (fonologicky /d/,/a/,/g/) z CDI je aktívny neurón reprezentujúci koncept [dog] a sieť je trénovaná na sekvencii ([dog],WB → /d/), ([dog],/d/ → /a/), ([dog],/a/ → /g/), ([dog],/g/ → WB). Trénovanie prebieha v dávkach tak, že zmeny váh sa akumulujú po každej foném, ale váhy sa naozaj zmenia až po prezentácii celého slova. Pre slovo mimo CDI, napr. „*ale*“ (/e/,/l/), bude trénovacia sekvencia ([ ],WB → /e/), ([ ],/e/ → /l/), ([ ],/l/ → WB), teda bez významu.

Sieť sme trénovali algoritmom spätného šírenia chyby v čase BPTT [14] s veľkosťou časového okna 4. Algoritmus BPTT spôsobí, že pri trénovaní na CDI slovách sa upravujú váhy medzi významovou a skrytou vrstvou (ktoré reprezentujú znalosť fonotaktiky špecifickú pre jednotlivé slová) aj váhy medzi vrstvou súčasnej fonémy a skrytou vrstvou (ktoré reprezentujú všeobecnú fonotaktiku), zatiaľ čo pre slová mimo CDI sa upravujú iba váhy reprezentujúce všeobecnú fonotaktiku.

Trénovanie trvalo 100 epoch, v každej epoche sa postupne predkladali všetky slová z trénovacej množiny v náhodnom poradí, medzi každými dvoma slovami sa resetoval kontext,[2] aby sa eliminoval vplyv predchádzajúceho slova. Parameter rýchlosti učenia lineárne klesal z 0.04 v prvej epoche na 0.01 v epoche 50 (potom ostal konštantný), aby sme zabránili osciláciám. Pre každú veľkosť skrytej vrstvy sme vytvorili skupinu 10 „participantov“ – inštancií modelu (H5, H10, H15, H20) – modely v rámci jednej skupiny boli inicializované inými náhodnými hodnotami váh a rôznymi stochasticky generovanými trénovacími množinami. Po každej epoche trénovania sme testovali schopnosť siete správne generovať sekvencie foném pre 268 CDI slov (bez úpravy váh počas testovania).

4 Výsledky

Z výsledkov simulácií pre všetky skupiny „participantov“ sme skonštruovali maticu *participant* x *epoch* x *data*, kde *data* boli tvorené skutočne vygenerovanými sekvenciami foném pre 268 CDI slov. Z tejto matice sme pre každého participanta vypočítali epochu osvojenia každého slova (*age of acquisition*, AofA) ako najskoršiu takú epochu, v ktorej participant vygeneroval pre daný význam

[1] Databáza [2] obsahuje v skutočnosti 4086 slov, ale 1508 z nich je takých zriedkavých, že sú uvedené s nulovou frekvenciou; tieto sme vylúčili.

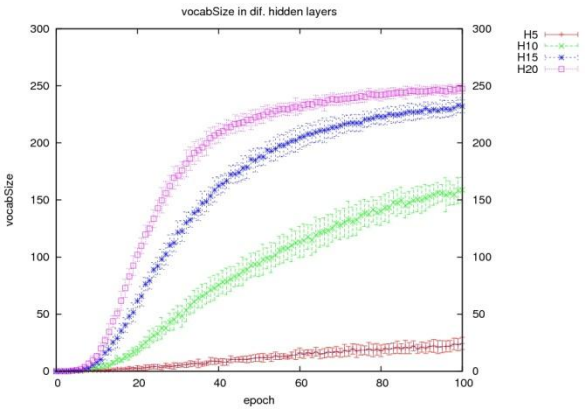
[2] Kontext sa resetoval desiatimi prechodmi siete „naprázdno“ – na vstupe ([ ], WB).

foneticky správne slovo. Slová, ktorá dokázal participant v danej epoche správne vygenerovať, tvoria jeho aktuálny lexikón. Zaujímala nás veľkosť lexikónu a ďalšie charakteristiky, napr. priemerná frekvencia výskytu slov z lexikónu v CELEX korpuse (WF), fonotaktická pravdepodobnosť – priemerná frekvencia prechodov medzi hláskami (biphone frequency BF)[3], priemerný počet fonologických susedov slova (ND) a priemerná dĺžka osvojených slov.

Zistili sme signifikantný efekt dĺžky slova a BF – dlhšie slová a slová zložené zo zriedkavejších prechodov medzi hláskami (s nižším BF) boli osvojované neskôr (mali vyššie AofA).

Ďalej sme zistili, že siete s menším počtom skrytých neurónov si osvojili menšie lexikóny a pomalšie (obr. 2).

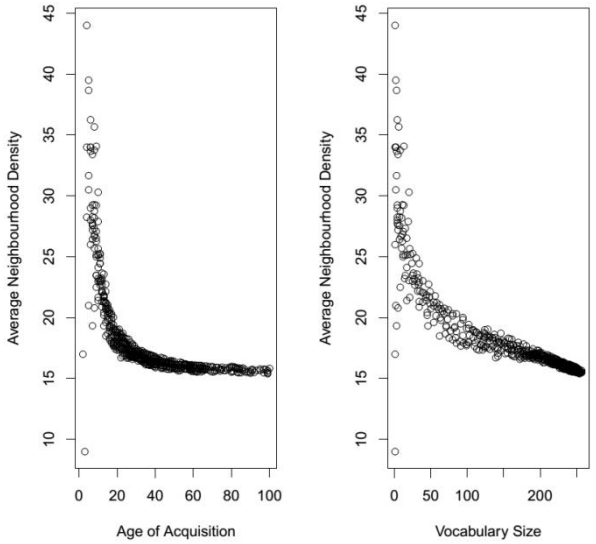
Zistili sme súvislosť medzi ND a AofA ako aj ND a veľkosťou lexikónu: neskôr boli osvojované slová s menším počtom susedov (obr. 3 vľavo), a malé lexikóny pozostávali zo slov s vyšším ND (obr. 3 vpravo).



Obr. 2. Závislosť počtu osvojených slov od veľkosti skrytej vrstvy v priebehu učenia. Každá krivka reprezentuje priemer z 10 modelov s rovnakou veľkosťou skrytej vrstvy (a smerodajnú odchýlku).

4.1 Porovnanie s empirickými dátami

Vzťah medzi priemerným ND lexikónu a jeho veľkosťou v našich simuláciách zodpovedá efektu nájdenému v produktívnych lexikónoch u anglických [8], francúzskych [11] a dánskych [9] detí (obr. 4). Základný trend od vysokej priemernej fonologickej susednosti po nižšiu so vzrastajúcou veľkosťou lexikónu je rovnaký u všetkých troch skupín detí ako aj v našom modeli, a tiež trend od veľkej variability ND po nižšiu. Celková variabilita v dátach vyprodukovaných naším modelom je nižšia ako u empirických dát. To môže byť spôsobené tým, že deti majú oveľa väčšiu pamäťovú kapacitu ako model, a tiež tým, že zatiaľ čo u detí zodpovedá každý bod v grafe inému dieťaťu, v grafe pre model pochádzajú všetky body z 10 inštancií (skupina H20) v rôznych epochách vývinu.



Obr. 3. Súvislosť veľkosti fonologického susedstva slova (Average Neighbourhood Density) s epochou osvojenia slova (Age of Acquisition, vľavo) a veľkosťou lexikónu (Vocabulary Size, vpravo).

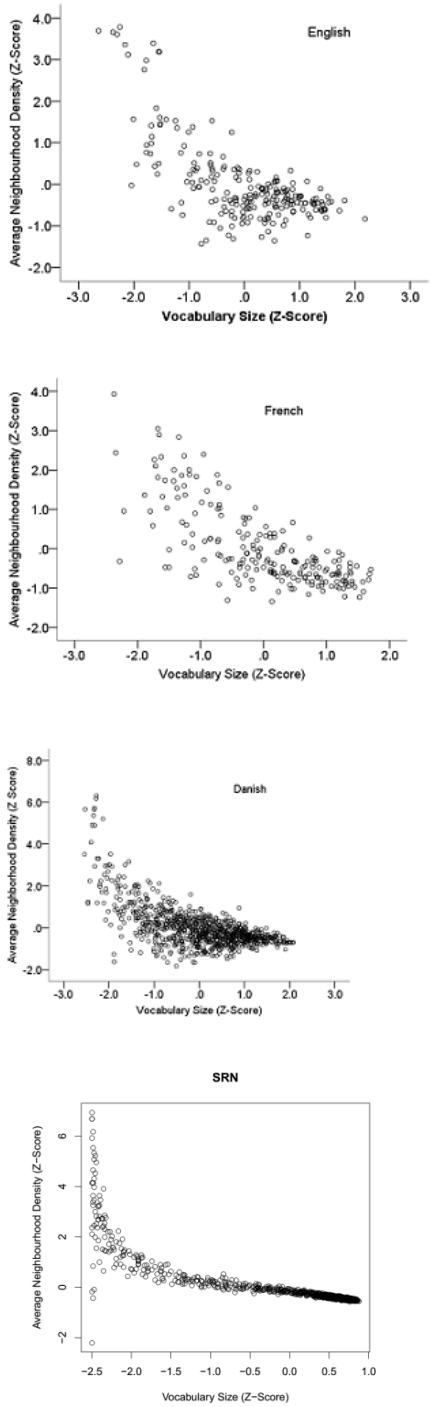
5 Vysvetlenie efektu fonologickej susednosti

Hierarchická regresná analýza výsledkov simulácií [13] ukázala, že efekt fonologickej susednosti nie je iba dôsledkom efektu dĺžky, frekvencie a fonotaktickej pravdepodobnosti slova, ale je nezávislý, keďže pretrváva aj po zahrnutí ostatných faktorov do modelu. Preto sa v tejto časti zameriame na vysvetlenie ND efektu.

5.1 Geometrická analýza siete

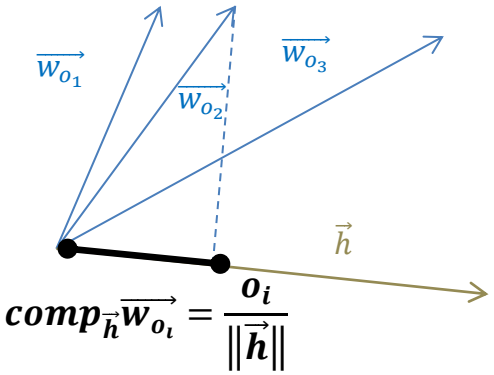
Výpočet šírenia aktivít v sieti má štandardnú geometrickú interpretáciu. Každý neurón na výstupnej vrstve reprezentuje jednu možnú nasledujúcu fonému a veľkosť jeho aktivity reprezentuje pravdepodobnosť tejto fonémy

[3] Táto miera vyjadruje, či je slovo zložené z typických, teda často sa vyskytujúcich prechodov medzi hláskami BF pre jedno slovo sa počíta ako suma relatívnych logaritmických frekvencií výskytov dvojíc hlások, z ktorých slovo pozostáva, v korpuse CELEX, vydelená dĺžkou slova. Priemerná BF pre celý lexikón je priemer BF pre jednotlivé slová.



Obr. 4. Súvislosť medzi veľkosťou fonologického susedstva slova (Average Neighbourhood Density) a veľkosťou lexikónu (Vocabulary Size) u anglických, francúzskych, dánskych detí (obr. z [8,11,9]) a v našom modeli.

v danom momente. Každý výstupný neurón o_i je spojený so skrytou vrstvou veľkosti n pomocou n spojení, tie spolu tvoria *vektor výstupných váh* $\overrightarrow{w_{o_i}}$. Keďže výstupné neuróny sú lineárne, aktivita každého z nich sa počíta ako skalárny súčin jeho vektora výstupných váh, a n-rozmerného vektora aktivít neurónov na skrytej vrstve, tzv. *skrytého stavového vektora* $\vec{h}$. Tieto skalárne súčiny možno interpretovať ako projekcie vektorov výstupných váh na skrytý stavový vektor (obr. 5) – fonéma, ktorej výstupný vektor váh má najdlhšiu projekciu na $\vec{h}$, sa stane víťazom a predstavuje predikovanú nasledujúcu fonému. V n-rozmernom stavovom priestore si môžeme predstaviť hyperguľu, ktorej povrch pretína skrytý stavový vektor aj vektory výstupných váh. Výstupné váhové vektory rozdelia povrch hypergule na oblasti bodov, ktoré sú k nim najbližšie (Voronoiho mozaika), pričom každá zodpovedá jednej nasledujúcej fonéme. Oblasť, do ktorej v danom momente ukazuje skrytý stavový vektor, určí, ktorá fonéma bude predikovaná. Počas generovania celej sekvencie foném daného slova skrytý vektor postupne mení polohu a jeho priesečníky s povrchom hypergule vytvárajú *trajektóriu* v stavovom priestore.



Obr. 5. Aktivita i-teho výstupného neurónu o_i je priamo úmerná skalárnej projekcii príslušného vektora výstupných váh $\overrightarrow{w_{o_i}}$ na skrytý stavový vektor $\vec{h}$.

Úloha siete počas učenia je upraviť výstupné váhové vektory (a tiež vektory váh medzi vstupnou a skrytou, a kontextovou a skrytou vrstvou, ktoré determinujú skrytý stavový vektor $\vec{h}$) tak, aby sieť správne predikovala víťaza pre každú pozíciu v každom slove z trénovacej množiny. Dôležitým aspektom trénovania je, že popri slovách, ktorých význam sieť pozná (cca 15% trénovacej množiny), je vo väčšine prípadov vystavená slovám, ktorým „nerozumie“, teda sa učí z nich len ako zo sekvencií foném bez významu (všetky neuróny na

vrstve Význam sú neaktívne). Pravidlo backpropagation modifikuje iba váhy spojení vychádzajúcich z neurónov, ktoré sú momentálne aktívne, preto pre slová bez významov sa spojenia z vrstvy Význam nemodifikujú. Ostatné spojenia v sieti preto reprezentujú hlavne všeobecnú znalosť fonológie (prechodové pravdepodobnosti medzi hláskami) a znalosť špecifická pre jednotlivé slová je uchovávaná iba v spojeniach medzi vrstvou Význam a skrytou vrstvou, ktoré sa trénujú menej ako 15% celkového času.

Úlohou týchto špecifických spojení je „vychýliť“ skrytý stavový vektor tak, aby v danom momente padol do oblasti nasledujúcej fonémy *pre daný význam*. Napríklad, všeobecná fonotaktická znalosť siete hovorí, že najpravdepodobnejšia prvá fonéma v anglických jednoslabičných slovách je /s/, avšak pokiaľ je na vstupe význam [dog], prvá fonéma má byť /d/, skrytý stavový vektor treba preto vychýliť tak, aby ukazoval do oblasti zodpovedajúcej hláske /d/ a nie /s/. Keďže význam na vstupe je konštantný počas generovania celého slova, vychýlenie stavového vektora spôsobené významom musí byť rovnaké pre každú pozíciu v danom slove. Sieť tak rieši problém minimalizácie chyby pre viaceré simultánne ohraničenia. Vyššie sme popísali aktivitu siete počas generovania slova ako trajektóriu v stavovom priestore, ktorý je výstupnými váhovými vektormi rozdelený na oblasti zodpovedajúce jednotlivým fonémam. Všeobecnú znalosť pravdepodobností prechodov medzi fonémami preto môžeme považovať za definovanie jednotlivých všeobecných trajektórií; úlohou spojení z významovej vrstvy je potom vychýliť danú trajektóriu tak, aby prechádzala požadovanými oblasťami správnymi pre fonémy daného slova.

Výchylka však spôsobí zmenu výstupu siete iba v prípade, ak daný bod trajektórie zmení vplyvom výchylky oblasť, do ktorej patrí. To umožňuje vytvoriť dostatočne malé výchylky, ktoré zachovávajú trajektóriu v sekvencii oblastí tam, kde je v súlade so všeobecnou fonológiou, a odchýlia ju do inej oblasti iba tam, kde treba. To, že sieť využíva malé analógové zmeny na dosiahnutie diskrétnej zmeny na výstupe SRN, je známe z literatúry a táto vlastnosť sa nazýva *tienenie* (angl. shading) [7].

Nájdenie takých váh spojení, ktoré spôsobia výchylku správnu pre každú pozíciu slova nemusí byť jednoduché a súvisí s veľkosťou skrytej vrstvy, ktorá definuje rozmer stavového priestoru. Ak je stavový priestor mnohorozmerný, je ľahšie v ňom reprezentovať viaceré ohraničenia súčasne: napríklad použitím ďalšej dimenzie/dimenzií na odchýlenie trajektórie v jednej pozícií tak, aby to neovplyvnilo výpočty na iných pozíciách. Tento problém je analogický rozmiestneniu k bodov v n-rozmernom priestore tak, aby poradie ich vzájomných vzdialeností zodpovedalo poradiu veľkostí ich vzájomných (externe zadaných) odlišností. Rieši ho technika *multidimensional scaling* [5] používaná v psychológii. Kruskal [5] ukázal, že čím vyššia je dimenzia n, tým lepšie je riešenie, a optimálne riešenie zachovávajúce správne poradie pre všetky odlišnosti existuje pre $n=k-1$ (tým netvrdíme, že backpropagation vždy nájde takéto riešenie).

Avšak aj bez ohľadu na veľkosť skrytej vrstvy úloha nájsť správnu výchylku je pre niektoré slová ľahšia ako pre iné. V prvom rade sú to slová, ktorých fonológia sa príliš neodlišuje od najpravdepodobnejších prechodov medzi fonémami, teda výchylka od „štandardnej“ trajektórie nemusí byť veľká. Ďalej sú to slová, ktoré znejú podobne ako iné slová, ktoré už sieť ovláda, pretože súčasná konfigurácia váh umožňuje vytvoriť výchylku trajektórie vedúcu k veľmi podobnej sekvencii foném, ktorú treba už len minimálne upraviť pre súčasné slovo. To, že sieť sa ľahšie učí slová s veľkým počtom fonologických susedov, je teda spôsobené tým, že ich požiadavky na výchylku trajektórie spôsobujú úpravu váh podobným smerom.

To zároveň vysvetľuje, prečo siete s nižším počtom skrytých neurónov majú problém osvojiť si slová s malým počtom fonologických susedov: Osvojenie takýchto slov znamená nutnosť reprezentovať veľa veľmi osobitých odchýliek od typických trajektórií zavedených všeobecnou fonológiou – výchyliek, ktoré sú každá iná. Ako sme uviedli vyššie, čím väčšia je dimenzionalita stavového priestoru, tým je to ľahšie: siete s malým počtom skrytých neurónov tak možno nemajú dostatočnú kapacitu na reprezentovanie všetkých osobitostí v slovách s malým počtom fonologických susedov.

6 Diskusia

V tomto článku sme prezentovali výpočtový model raného fonologického a lexikálneho vývinu zameraný na výskum efektu fonologického susedstva pomocou jednoduchej rekurentnej siete. Trénovacie dáta siete boli odvodené z reálnej vzorky jazyka, frekvencie slov a teda fonotaktiku sme prebrali z veľkého korpusu hovorenej angličtiny a niektoré slová boli spárované s významami v súlade s normami detského lexikálneho vývinu. Sieť trénovaná na týchto dátach vykazovala preferenciu pre slová s vyšším počtom fonologických susedov, ako aj preferenciu pre frekventovanejšie a kratšie slová a slová zložené z frekventovaných kombinácií hlások. Efekt fonologického susedstva je porovnateľný s rovnakým efektom u detí vo viacerých smeroch: je najsilnejší v čase, keď sú osvojované prvé asociácie medzi slovami a významami, a je výraznejší pri nižšej kapacite skrytej vrstvy v modeli resp. fonologickej pracovnej pamäti u detí.

Poďakovanie

Tento výskum vznikol v spolupráci s Alistairom Knottom (University of Otago), Stephanie Stokes a Jennifer Hay (obe z Canterbury University) a bol podporený grantami VEGA 1/0898/14 = 3700 a KEGA 076UK-4/2013.

Literatúra

[1] R. H. Baayen, R. Piepenbrock, H. van Rijn: The CELEX lexical database (CD-ROM). Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania, 1995.

[2] B. De Cara, U. Goswami: Statistical analysis of similarity relations among spoken words: Evidence for the special status of rimes in English. *Behavioural Research Methods and Instrumentation* 34/3 (2002) 416-423.

[3] J. Elman: Finding structure in time. *Cognitive Science* 14 (1990) 179-211.

[4] L. Fenson, P. Dale, J. S. Reznick, D. Thal, E. Bates, J. Hartung, S. Pethick, J. Reilly: Variability in early communicative development. *Monographs of the Society for Research in Child Development* 59/5 (1994) i-185.

[5] J. B. Kruskal: Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* 9/1 (1964) 1-27.

[6] J. Moyle, S. Stokes, T. Klee: Early language delay and specific language impairment. *Developmental Disabilities Research Reviews* 17 (2011) 160-169.

[7] D. Servan-Schreiber, A. Cleeremans, J. L. McClelland: Graded state machines: The representation of temporal contingencies in simple recurrent networks. *Machine Learning* 7/2-3 (1991) 161-193.

[8] S. Stokes: The impact of phonological neighbourhood density on typical and atypical emerging lexicons. *Journal of Child Language* (2013), doi:10.1017/S030500091300010X.

[9] S. Stokes, D. Bleses, H. Basb, C. Lambertsen: Statistical learning in emerging lexicons: The case of Danish. *Journal of Speech, Language, and Hearing Research* 55 (2012) 1265-73.

[10] S. Stokes, T. Klee: Factors that influence vocabulary development in two-year old children. *Journal of Child Psychology and Psychiatry* 50 (2009) 498-505.

[11] S. Stokes, S. Kern, C. dos Santos: Extended statistical learning as an account for slow vocabulary growth. *Journal of Child Language* 39/1 (2012) 105-129.

[12] H. L. Storkel: Do children acquire dense neighborhoods? An investigation of similarity neighborhoods in lexical acquisition. *Applied Psycholinguistics* 25 (2004) 201-221.

[13] M. Takáč, A. Knott, S. Stokes, J. Hay: A simple recurrent network model of neighbourhood density effects in vocabulary development, pripravované.

[14] P. J. Werbos: Backpropagation through time: what it does and how to do it. *Proceedings of the IEEE* 78/10 (2002) 1550-1560.