

# Učenie nesusedných závislostí pomocou rekurentných neurónových sietí \*

Igor Farkaš a Ján Švantner

Katedra aplikovanej informatiky FMFI  
Univerzita Komenského  
Mlynská dolina, 84248 Bratislava  
farkas,svantner@ii.fmph.uniba.sk

## Abstrakt

Ľudia dokážu pri osvojovaní si a používaní jazyka využívať štatistické závislosti medzi elementami jazyka (slabiky, slová), pričom táto ich schopnosť sa netýka len susedných, ale aj nesusedných elementov, čo je zo štatistického pohľadu zložitejšia úloha. Zatiaľ čo schopnosť segmentovať slová zo súvisleho sekvenčného vstupu zrejme možno vysvetliť pomocou štatistických mechanizmov, v prípade schopnosti generalizovať (na základe odhalenia nesusedných syntaktických závislostí) vzniká otázka, či nie je potrebný iný, algebraický výpočtový mechanizmus. V príspevku modelujeme tento problém pomocou dvoch rekurentných neurónových sietí, Elmanovej siete a siete s echo stavmi, a testujeme tzv. hypotézu variability, podľa ktorej štatistické učenie umožňuje za istých podmienok vysvetliť aj schopnosť generalizácie. Ukážeme, že správanie modelov podporuje hypotézu variability.

## 1 Úvod

Predmetom dlhodobého záujmu je spôsob a mechanizmy, pomocou ktorých si človek osvojuje materinský jazyk. Viaceré teoretické i empirické štúdie poukazujú na to, že tento proces osvojovania si jazyka vyžaduje nielen existenciu človeka v jazykovom prostredí, ale aj nejaké vrodené dispozície pre spracovanie sekvenčných vstupov (Chomsky, 1957, Seidenberg, 1997). V súčasnosti neexistuje konsenzus o presnom fungovaní tohto procesu (debata „vrodené verus získané“), no ľahkosť, s akou si dieťa osvojuje jazyk navodzuje dojem, že tento proces je inštinktívny a že si nevyžaduje žiadnu zvláštnu námahu (Pinker, 1994). Súčasne však viaceré psychologicky plauzibilné konekcionistické modely naznačujú, že rôzne aspekty jazyka (na lexikálnej či vetnej úrovni) sú naučiteľné na základe štatistických závislostí (pozri prehľad v Rohde, Plaut, 2003; Farkaš, 2005a). Štatistický pohľad je konzistentný s predpokladom, že hlavným cieľom dieťaťa je porozumieť tomu, čo sa hovorí, a že samotné učenie sa jazyka je skôr iba sprievodným javom.

Štatistické učenie je dôležitým mechanizmom (aj keď možno nie postačujúcim) pri osvojovaní si a používaní jazyka. Je totižto známe, že pri počúvaní reči ľudia dokážu na rôznych úrovniach využívať distribučné charakteristiky elementov jazyka, ktoré im umožnia odhaliť štrukturálne závislosti v jazyku. Napríklad, štatistické vlastnosti kontextov slov sú relatívne dobrým prediktorom gramatických kategórií slov, obzvlášť v jazykoch s pomerne pevným slovosledom (Redington, Chater, Finch, 1998). Podobne na sublexikálnej úrovni dokážu ľudia (dojčatá, malé deti i dospelí) pomocou prechodových pravdepodobností medzi susednými slabikami detekovať slová v súvislej reči (Saffran, Newport, Aslin, 1996; Saffran a spol., 1997). Taktiež je známe, že štatistické učenie hrá rolu i pri osvojovaní si syntaxe a morfológie jazyka (Mintz, Newport, Bever, 2002; Saffran, 2002). Súdiac podľa empirických poznatkov možno konštatovať, že ľudia využívajú štatistické informácie nevedome, ako súčasť implicitného učenia (Cleeremans, Destrebecqz, Boyer, 1998).

Okrem susedných závislostí sa však v prirodzenom jazyku vyskytujú aj závislosti medzi vzdialenejšími elementami vety (nesusedné závislosti), ako napr. v angličtine zhoda v čísle medzi podstatným menom a slovesom (*boys in the street are naughty*) alebo zhoda medzi pomocnými slovesami a sufixami (*is sleeping, has arrived*).<sup>1</sup> Táto schopnosť človeka bola tiež testovaná na umelých jazykoch<sup>2</sup>, no závery týchto experimentov nie sú jednoznačné (Newport, Aslin, 2004; Onnis a spol., 2004; Pena a spol., 2002). Experimenty s učením umelých jazykov poukázali na to, že človek nedokáže zovšobecňovať (t.j. napr. predikovať príponu *-ing*, ak po *is* bol na vstupe nový koreň *eat*) len na základe štatistickej informácie zo súvislých akustických sekvenčných vstupov, pokiaľ spolu s nimi nie sú poskytované aj nejaké perceptuálne „značky“ (cues) ako

<sup>1</sup> Samozrejme to platí aj pre morfológicky bohatšiu slovenčinu (napr. *knihy sú zaprášené*).

<sup>2</sup> Ide o bežnú metodológiu v psycholingvistiky, pri ktorej síce pracujeme so zjednodušenými jazykmi, no experiment máme pod kontrolou (môžeme systematicky manipulovať nezávislé premenné), a výsledky majú relevantnú výpovednú hodnotu.

\* Práca bola podporená agentúrou VEGA.

pauzy medzi slovami alebo fonologické charakteristiky. Tento poznatok vyústil v hypotézu, podľa ktorej štatistické informácie samotné nepostačujú na to, aby sa v súvislej reči dali odhaliť vzťahy, ktoré sa sa dajú vyjadriť pomocou gramatických pravidiel. Pena a spol. (2002) robili experimenty s dospelými ľuďmi, ktorí sa učili umelý jazyk pozostávajúci z pospájaných trojslabičných slov s rámcom  $A_iX_kB_i$  (kde  $A_i$  jednoznačne predikuje  $B_i$  bez ohľadu na „výplň“  $X_k$ ) a na základe výsledkov argumentovali, že segmentácia slov a generalizácia sú separátne procesy zabezpečované rôznymi mechanizmami: štatistické výpočty sa využívajú pri segmentácii, no tie sú nezávislé od algebraických výpočtov na báze pravidiel, ktoré vysvetľujú schopnosť generalizovať pri nesusedných (syntaktických) závislostiach.

Avšak, ako promptne argumentovala Gómez (2002), neschopnosť ľudí generalizovať mohla vzniknúť kvôli nízkej variabilite výplne ( $|X|$ ), pretože v jej experimentoch s vyššou variabilitou ( $|X| = 24$ ) boli ľudia schopní generalizovať. Onnis a spol. (2003, 2004) zopakovali jej výsledky a zjasnili hypotézu variability postulovaním, že generalizácia vzniká pri oboch extrémoch – pri nulovej a pri veľkej variabilite: Keď veľká variabilita výplne naruší susedné závislosti, človek začne hľadať alternatívne zdroje informácie (ako napr. nesusedné závislosti) v snahe predikovať ďalšie elementy vety. Pri nulovej variabilite výplne zase vzniká opačný efekt, lebo ten istý element  $X$  zdieľa rovnaké kontextuálne rámce (napr. *don't eat it, he's eat-ing*). Onnis a spol. (2004) na dátach od Pena a spol. ukázali, že pri dostatočne vysokej variabilite výplne môže sledovanie nesusedných závislostí viesť k schopnosti človeka simultánne segmentovať slová i generalizovať pri nových výplniach.<sup>3</sup> K úspešnému naučeniu pritom došlo len na základe štatistickej analýzy súvislého sekvenčného vstupu, t.j. bez akýchkoľvek perceptuálnych charakteristík.

V tomto príspevku modelujeme hypotézu variability na dátach umelého jazyka pomocou dvoch modelov rekurentných neurónových sietí – štandardnej Elmanovej jednoduchej rekurentnej siete (Elman, 1990), a pomocou siete s echo stavmi (Jaeger, 2001).

## 2 Konekcionistické modelovanie

### 2.1 Špecifikácia vstupných dát

Ako vstupy budeme používať 3 sady slov ( $S_1, S_2, S_3$ ), pričom každá sada obsahuje trojslabičné slová ( $M=3$ ) z

<sup>3</sup> Treba podotknúť, že už samotná segmentácia slov zo súvislého sekvenčného vstupu je obtiažna, lebo veľká variabilita výplne vedie k segmentácii vnútri samotných slov (Saffran a spol., 1996).

umelého jazyka. Tieto (gramatické) slová (GW) sú tvaru  $A_iX_kB_i$ . Budeme používať tri rámce  $A_iB_i$  (s rovnakou pravdepodobnosťou) spoločne pre všetky 3 sady: *ba\_te, gu\_do* a *pi\_ra*. Jednotlivé sady vstupov (jazyky) sa od seba líšia počtom slabík, ktoré sú použité ako  $X$  (variabilita  $X$ ):  $X_1 = \{di\}$ ,  $X_2 = \{di, ku, to, pa\}$ ,  $X_3 = \{be, bi, bo, bu, ta, ti, to, tu, ga, \dots\}$ , odkiaľ vidieť, že  $|X_1| = 1$  (nulová variabilita),  $|X_2| = 4$  (malá) a  $|X_3| = 29$  (veľká). Slabiky, ktoré sú v pozícii  $X$ , sa nevyskytujú v  $A$  ani v  $B$ , taktiež množiny  $A$  a  $B$  sú disjunktné. Čo sa týka rôznych prechodových pravdepodobností<sup>4</sup>, tak potom dostávame:  $P(X_k|A_i) = 1/|X|$ ,  $P(B_i|A_i) = 1$ ,  $P(B_i|X_k) = 0.33$ , a medzi slovami  $P(A_i|B_i) = 0.5$  (slová sa nemohli hneď po sebe opakovať). Vstupná postupnosť je prezentovaná po slabikách, tie sú kódované lokalisticky (one-hot), t.j. pre každú slabiku má len jeden prvok vstupného vektora rovný 1, ostatné sú 0. Tento spôsob kódovania určuje architektúru siete, konkrétne počet neurónov na vstupnej a výstupnej vrstve (keďže aj výstup kódujeme lokalisticky).

Ešte vysvetlíme niekoľko pojmov (podľa Pena a spol., 2002), ktoré budeme používať pri opise experimentov. Budeme ich ilustrovať na vstupnej sade  $S_1 = \{badite, gudidido, pidira\}$ . Vstup siete pri tréningu tvorí postupnosť slabík z  $A \cup B \cup X$ , ktorá vznikne zretazením slov príslušnej vstupnej sady v náhodnom poradí (napr. *gudidobaditepidiragudidido...*). Pod pojmom *part-word* (PW) budeme rozumieť slovo, ktoré vznikne spojením 3. slabiky jedného slova a 1. a 2. nasledujúceho slova, alebo spojením 2. a 3. slabiky jedného slova a 1. slabiky nasledujúceho slova. Napr. v reťazci

*gudidobaditepidiragudidopidirapidirabadite*

sú PW zvýraznené hrubo a každé druhé slovo jazyka je nadčiarknuté. Slová PW predstavujú ostatné možnosti delenia vstupnej postupnosti slabík na slová pevnej, vopred známej dĺžky (v našom prípade 3). Budeme ich teda využívať pri testovaní segmentácie slov (porovnávaním so slovami). Takto môžeme zistiť, nakoľko testovaná sieť preferuje delenie vstupnej postupnosti zodpovedajúce slovám GW pred iným delením (t.j. prekračujúcim hranice slov).

Pod pojmom *rule-word* (RW) budeme označovať slová, ktoré majú tvar  $A_iA_jB_i$  alebo  $A_iB_jB_i$ . Príkladmi RW pre  $S_1$  sú *bagute* alebo *pidora*. Slová RW teda ako výplň obsahujú slabiky, ktoré sa vyskytli počas tréningu, ale nikdy v pozícii výplne. Slabiky RW slova sú usporiadané v poradí, v akom sa v pôvodnej tréningovej postupnosti nevyskytli. Prechodové pravdepodobnosti  $P(A_i|A_i) = P(B_i|A_i) = 0$ , preto jediná väzba je medzi 1. a 3. slabikou. Keďže slovo RW je vytvorené podľa abstraktného,

<sup>4</sup>  $P(X|Y)$  znamená podmienenú pravdepodobnosť, že nastane udalosť  $X$ , ak nastala udalosť  $Y$ .

algebraického pravidla, budeme pomocou neho testovať schopnosť siete aj zovšeobecňovať (t.j. nielen schopnosť identifikovať slová, ale aj odhaliť nesusedné štrukturálne závislosti), a to porovnávaním so slovami PW. Testovanú hypotézu možno teda formulovať takto: ak učiaci systém počíta susedné pravdepodobnosti, mal by preferovať slová PW, a to aspoň v prípade malej  $|X|$ . lebo ich vnútorné prechodové sú väčšie. A naopak, ak počíta nesusedné závislosti, tak sa bude spoliehať na tie štatisticky najspoľahlivejšie, ktorými sú  $P(B_i|A_i)$ , vďaka čomu bude vedieť správne segmentovať slová (vrátane slov RW).

### 3.1 Špecifikácia modelov

Pri oboch modeloch bol počet vstupných aj výstupných neurónov pre sadu S1 rovný 7, pre S2 10 a pre S3 35. Oba modely mali za úlohu predikovať nasledujúcu slabiku. Pri modeli SRN sme počet neurónov skrytej vrstvy zvolili 10. Ako aktivačnú funkciu neurónov v skrytej vrstve sme používali unipolárnu sigmoidu (so strmosťou 2), a v prípade výstupnej vrstvy aktivačnú funkciu typu softmax. Vstupné i výstupné váhy boli náhodne inicializované z intervalu  $[-0.5, 0.5]$ . Na adaptáciu váh bol použitý štandardný algoritmus spätného šírenia chyby (Rumelhart, Hinton, Williams, 1986) s rýchlosťou učenia 0.1. Použili sme 20 neurónov v skrytej vrstve. V prípade ESN boli vstupné váhy náhodne nastavené na -0.1 alebo 0.1, s rovnakou pravdepodobnosťou, výstupné váhy boli inicializované v intervale  $[-0.5, 0.5]$ . Použili sme 50 neurónov v skrytej vrstve. Matica kontextových váh (určujúca dynamický rezervoár stavov siete) bola riedka (80% váh nulových) a ostatné hodnoty boli inicializované na malé hodnoty tak, aby bola splnená podmienka pre existenciu echo stavov, t.j. aby spektrálny radius matice  $|\lambda_{\max}| < 1$ , kde  $\lambda_{\max}$  je najväčšie vlastné číslo matice (Jaeger, 2002). Zvolili sme  $|\lambda_{\max}| = 0.8$ . Adaptované boli len výstupné váhy, pomocou štandardného delta pravidla pre perceptron, s rýchlosťou učenia 0.1.

V psycholingvistických experimentoch sa mali subjekty po počutí dvoch testovaných slov rozhodnúť, ktoré im viac pripomínalo slovo GW. Pri vyhodnocovaní správania natrénovanej siete sme preto museli nájsť vhodné kritérium, pomocou ktorého sieť rozhodne, ktoré zo slov sa jej viac „páči“. Architektúra našej siete neumožňuje zadať na vstup siete dve slová naraz a priamo získať výsledok, navyše musíme jednotlivé slová zadávať postupne po slabikách v priebehu viacerých taktov. Preto sme navrhli nasledovný spôsob vyhodnocovania:

- vyberieme rovnaký počet ( $N = 20$ ) slov a „neslov“
- pre každé slovo zistíme jeho ohodnotenie

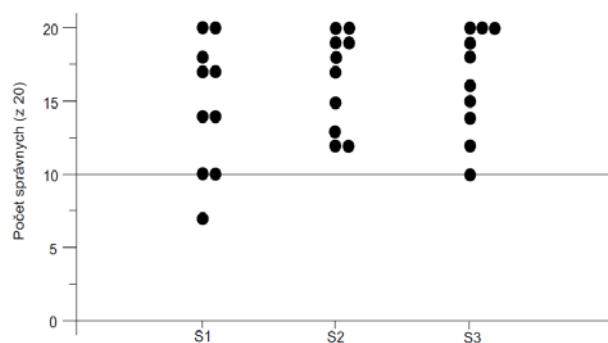
- vyberieme  $N$  slov s najvyšším ohodnotením a zistíme, aká časť z nich patrí do prvej množiny (slová GW resp. RW), a koľko do druhej „neslová“ (PW).

Ohodnotenie slova získame tak, že:

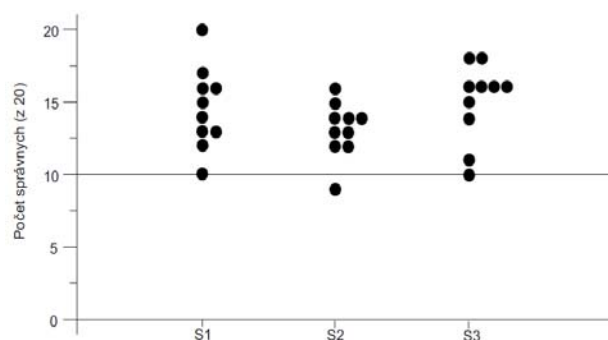
- zresetujeme kontextovú vrstvu siete (nastavením aktivity neurónov na 0), čím zabezpečíme, že sieť bude izolovaná od predchádzajúcich vstupov,
- postupne zadávame na vstup prvé dve slabiky a zaznamenáme výstupy neurónu ( $u_i$ ) odpovedajúceho nasledujúcej slabike v každom takte,
- ohodnotenie slova definujeme ako  $u = u_1 + u_2$ .

### 3.2 Výsledky

Pri vyhodnocovaní siete na segmentáciu slov sme porovnávali testované slová GW a PW. Výsledky experimentu so sieťou SRN sú na obr. 1. Priemerný výsledok pre S1 bol 14.7 z 20-tich (73.5%), pre S2 16.5 (82.5%) a pre S3 16.4 (82%). Ako vidno, sieť SRN úspešne zvláda segmentovanie slov. Podobná situácia nastala pri modeli ESN na obr. 2. Priemerný výsledok pre S1 bol 14.6 z 20-tich (73%), pre S2 13.2 (65%) a pre S3 15.0 (75%).



**Obr. 1.** Výsledky segmentácie slov pomocou SRN. Sieť preferuje slová GW pred PW nezávisle od variability výplne. Vodorovná čiara pri hodnote 10 znamená výsledok pri náhodnom výbere slov z dvojíc.

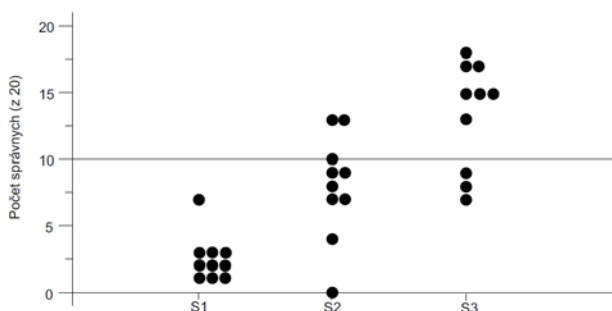


**Obr. 2.** Výsledky segmentácie slov pomocou ESN. Sieť preferuje slová GW pred slovami PW nezávisle od variability výplne.

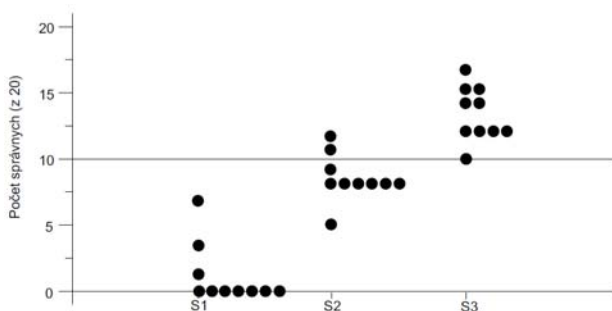
Pri testovaní generalizácie siete sme porovnávali jej predikcie na slovách RW (slová) a PW („neslová“). Výsledky pre model SRN sú na obr. 3. Priemerný výsledok pre S1 bol 2.5 z 20-tich (12.5%), pre S2 8.0 (40%) a pre S3 13.4 (67%). Nami získané výsledky sa vo veľkej miere zhodujú s výsledkami experimentu 2 v práci Onnis a spol. (2004). Spomínaný experiment sa týkal generalizácie, kde vstupná sekvencia bola súvislá (teda hranice slov neboli označené pauzami). Pokus robili pri nízkej ( $|X| = 3$ ) a vysokej ( $|X| = 24$ ) variabilite (v našom pokuse nízka ( $|X| = 4$ ) a vysoká ( $|X| = 29$ )). Výsledné úspešnosti preferencie RW pred PW v ich experimente boli 41.9% pre nízku variabilitu (u nás S2 a 40%) a 67% pre vysokú variabilitu (u nás S3 a 67%). V ich experimente nie je uvedená úspešnosť pre nulovú variabilitu, takže túto hodnotu nemôžeme porovnať.

Podobné výsledky sme dostali s modelom ESN na na obr. 4. Priemerný výsledok pre S1 bol 1.1 z 20-tich (5.5%), pre S2 8.5 (42.5%) a pre S3 13.2 (66%).<sup>5</sup>

Treba dodať, že v oboch prípadoch naše výsledky nie sú konzistentné s hypotézou variability pri nulovej variabilite (preferencie slov RW sú veľmi malé). Avšak, podľa Onnis a spol. (2003, 2004), bola vysoká preferencia RW u subjektov preukázaná iba v prípade segmentovaného vstupu.



**Obr. 3.** Výsledky testu generalizácie v modeli SRN. Sieť generalizuje do istej miery pri veľkej variabilite výplne (S3).



**Obr. 4.** Výsledky generalizácie v modeli ESN. Sieť generalizuje do istej miery pri veľkej variabilite výplne.

<sup>5</sup> Pôvodné experimenty pri použití malého počtu skrytých neurónov v ESN boli neúspešné (Černák, 2005).

V práci Farkaš (2005b) bol pomocou modelu SRN testovaný aj realistickejší prípad dlhších závislostí, konkrétne na prípade dvojslabičnej výplne. Slová boli generované podľa rámca  $A_i X_k Y_l B_i$ , pričom slabiky výplne  $X_k$  a  $Y_l$  boli nezávislé. V takomto prípade sa podarilo dosiahnuť istú úroveň generalizácie pri väčšej variabilite len pri použití výkonnejšieho učiaceho algoritmu v reálnom čase (tzv. real-time recurrent learning; Williams, Zipser, 1989).

#### 4. Diskusia

Štatistické učenie závislostí medzi elementami sekvencií je automatický proces, ktorý ľudia bežne využívajú pri spracovaní časových štruktúr. Skoršie práce poukazovali na to, že s tým súvisiace výpočty sa týkajú susedných elementov, ale podľa novších prác sa takéto výpočty týkajú aj nesusedných závislostí. Táto úloha sa javí byť ťažšia, práve možno kvôli tendencii človeka opierať sa o prechodové pravdepodobnosti medzi susednými elementami. Okrem toho, pri nesusedných závislostiach čelí človek kombinatorickému problému, pretože počet možných nesusedných pravdepodobností, ktoré možno sledovať, rastie exponenciálne s dĺžkou výplne. Preto možno predpokladať, že o výpočty medzi vzdialenými elementami sa možno opierať iba za istých podmienok.

Pri hľadaní takýchto podmienok sa dospelo v skorších práce k hypotéze, že učenie nesusedných závislostí sa dá dosiahnuť, len ak sú pridané dotatočné značky. Pena a spol. dospeli k záveru, že sú nutné (sumliminálne) pauzy medzi slovami. Newport a Aslin (2004) konštatovali, že sú nutné fonologické charakteristiky. Avšak podľa Onnis a spol. (2004) k učeniu nesusedných závislostí môže dôjsť aj v súvislom sekvenčnom vstupe bez týchto značiek, a to za predpokladu, že variabilita výplne je dostatočne veľká. Naše experimenty potvrdili túto výpočtovú možnosť v prípade sekvenčných učiacich sa modelov neurónových sietí, ktoré sa opierajú iba o poradie elementov na vstupe a principiálne pritom využívajú iba štatistické výpočty.

Sledovanie nesusedných závislostí je vo všeobecnosti obtiažna lingvistická úloha ako pre ľudí tak aj pre neurónové siete, ak ide napr. o súvetia s vnorenými vetami (Cleeremans, McClelland, 1991; Christiansen, Chater, 1999). O rekurentných neurónových sieťach bolo síce aj ukázané, že sa dokážu naučiť bezkontextové jazyky napr. typu  $a^n b^n$  (Rodriguez, Wiles, Elman, 1999; Bodén, Wiles, 2000), no učiace procesy v týchto prípadoch boli študované na vyššej, abstraktnejšej úrovni, pri uvažovaní iba malého počtu symbolov (dvoch, v prípade jazyka  $a^n b^n$ ). Toto je iná situácia, než nami adresovaná úroveň slabík.

Záverom možno povedať, že sledovanie nesusedných vzdialeností je pravdepodobne dôležitá lingvistická ľudská schopnosť, ktorej podstatu len začíname

odhaľovať. Vzniká tu viacero otázok, jednou z ktorých je tá, či susedné a nesusedné závislosti predstavujú separátne učiace procesy alebo sa za nimi skrýva ten istý všeobecný mechanizmus. Empirické výsledky Gómez (2002) a Onnis a spol. (2004) ako aj naše simulácie podporujú tvrdenie, že spoločný učiaci mechanizmus môže dokázať realizovať rôzne štatistické výpočty hľadájúc pritom najspoľahlivejšie zdroje informácie. Táto hypotéza je konzistentná s hypotézou o „redukcii neistoty“ (Gibson, 1991) alebo princípu jednoduchosti (Chater, 1996), ktoré postulujú, že učiaci sa systém má tendenciu vybrať najjednoduchšiu možnú hypotézu o dostupných dátach, a to hľadaním ich invariantných vzorov. Keď sú prechodové pravdepodobnosti vysoké, susedné elementy sú vnímané ako invariantné. Keď však veľká variabilita naruší susedné pravdepodobnosti, učiaci sa systém sa preladí na alternatívne zdroje invariantnosti, potenciálne existujúce medzi nesusednými elementami.

## Literatúra

- [1] Bodén M., Wiles J.: Context-free and context-sensitive dynamics in recurrent neural networks. *Connection Science* 12 (2000) 197-210.
- [2] Chater, N.: Reconciling simplicity and likelihood principles in perceptual organization. *Psychological Review* 103 (1996) 566-581.
- [3] Chomsky N.: *Syntactic Structures*. Mouton, Hague, 1957.
- [4] Christiansen M., Chater N.: Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science* 23 (1999) 157-205.
- [5] Cleeremans, A. *Mechanisms of Implicit Learning*. MIT Press, Cambridge, MA, 1993.
- [6] Cleeremans A., Destrebecqz A., Boyer M.: Implicit learning: news from the front. *Trends in Cognitive Sciences* 2 (1998) 406-416.
- [7] Cleeremans A., McClelland J.: Learning the structure of event sequences. *Journal of Experimental Psychology: General* 120 (1991) 235-253.
- [8] Černák M.: Učenie nesusedných závislostí pomocou rekurentných neurónových sietí. *Diplomová práca*. FMFI, Univerzita Komenského, 2005.
- [9] Elman J.: Finding structure in time. *Cognitive Science* 14 (1990) 179-211.
- [10] Farkaš I.: Konekcionalistické modelovanie jazyka. In *Jazyk a kognícia* (Rybár J., Kvasnička V., Farkaš I., zost.), Kalligram, Bratislava, 2005a:262-305.
- [11] Farkaš I.: Learning non-adjacent dependencies with a simple recurrent network. *Pracovný text*. FMFI, Univerzita Komenského, 2005b.
- [12] Gibson E.: *An Odyssey in Learning and Perception*. MIT Press, Cambridge, MA, 1991.
- [13] Gómez R.: Variability and detection of invariant structure. *Psychological Science* 13 (2002) 431-436.
- [14] Jaeger H.: Short-term memory in echo state networks. *GMD Report 152*, German National Research Center for Information Technology, 2002.
- [15] Jaeger H.: The „echo state“ approach to analysing and training recurrent neural networks. *GMD Report 148*, German National Research Center for Information Technology, 2001.
- [16] Mintz T., Newport E., Bever T.: The distributional structure of grammatical categories in speech to young children. *Cognitive Science* 26 (2002) 393-424.
- [17] Newport E., Aslin R.: Learning at a distance I. Statistical learning of nonadjacent dependencies. *Cognitive Psychology* 48 (2004) 127-162.
- [18] Onnis L., Christiansen M., Chater N., Gómez R.: Reduction of uncertainty in human sequential learning: Evidence from artificial grammar learning. *Proc. of the 25th Annual Conf. of the Cog. Sci. Soc.*, 887-891, 2003.
- [19] Onnis L., Monaghan P., Christiansen M., Chater N.: Variability is the spice of learning, and a crucial ingredient for detecting and generalizing in nonadjacent dependencies. *Proc. of the 26th Annual Conf. of the Cog. Sci. Soc.*, 1047-1052, 2004.
- [20] Pena M., Bonatti L., Nespor M., Mehler J.: Signal-driven computations in speech processing. *Science* 298 (2002) 604-607.
- [21] Pinker S.: *Language instinct*. Morrow, NY, 1994.
- [22] Redington M., Chater N., Finch S.: Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive Science* 22 (1998) 425-470.
- [23] Rodriguez P., Wiles J., Elman J.: A recurrent neural network that learns to count. *Connection Science* 11 (1999) 5-40.
- [24] Rohde D., Plaut D.: Connectionist models of language processing. *Cog. Studies* 10 (2003) 10-28.
- [25] Rumelhart D., Hinton E., Williams J.: Learning representations by back-propagating errors. *Nature* 323 (1986) 533-536.
- [26] Saffran J., Aslin R., Newport E.: Statistical learning by 8-month old infants. *Science* 274 (1996) 1926-1928.
- [27] Saffran J., Newport E., Aslin R., Tunick R., Barrueco S.: Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science* 8 (1997) 101-105.
- [28] Seidenberg M.: Language acquisition and use: learning and applying probabilistic constraints. *Science* 275 (1997) 1599-1603.
- [29] Williams R., Zipser D.: A learning algorithm for continually running fully recurrent neural networks. *Neural Computation* 1 (1989) 270-280.