A New Theoretical Model for Word Finding Difficulties in Aphasic Patients

Marjolein van Egmond, Lizet van Ewijk and Sergey Avrutin

Universiteit Utrecht

UiL-OTS, Universiteit Utrecht, The Netherlands

Utrecht Institute of Linguistics OTS

1. Introduction

Non-fluent aphasia in a nutshell:
- Acquired language disorder
- Difficulties with closed class words
- **Word finding problems for open class (content) words**

Focus for the present study: content words only.

Dominant explanation: aphasic speakers suffer from reduced processing capacities. Lexical selection fails if words require too much effort to process.

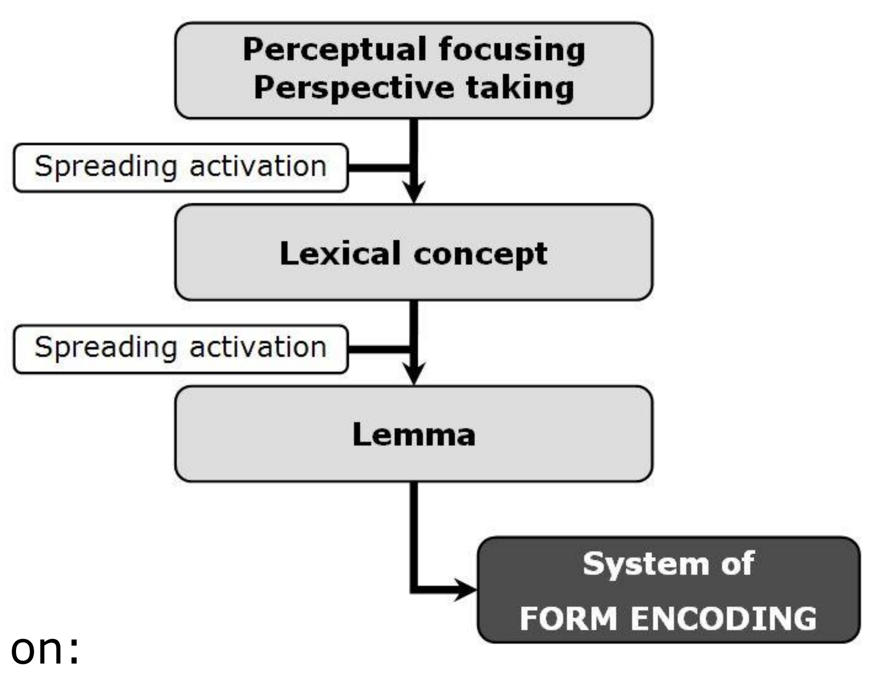
Question: which words are most difficult to process, and why these words?

Lower processing capacities imply lower processable entropy in aphasic speakers. Therefore, it is hypothesized that entropy is unconsciously reduced in the most efficient way to maintain as much of the lexicon as possible. The question is: **how is entropy most efficiently reduced?**

2. Lexical Retrieval

Words are retrieved from the lexicon through the following system:

Lexical selection starts by focusing on the concept that a speaker wants to talk about. Activation spreads to the lexical concept that the speaker wants to express, and from there to the specific lemma he wants to use. The lemma with the highest activation is selected for lexical retrieval.



The ease of lexical retrieval depends on:
- Base level of activation (mainly determined by recent activity)
- The strength of lexical connections
- **The number of lexical connections**

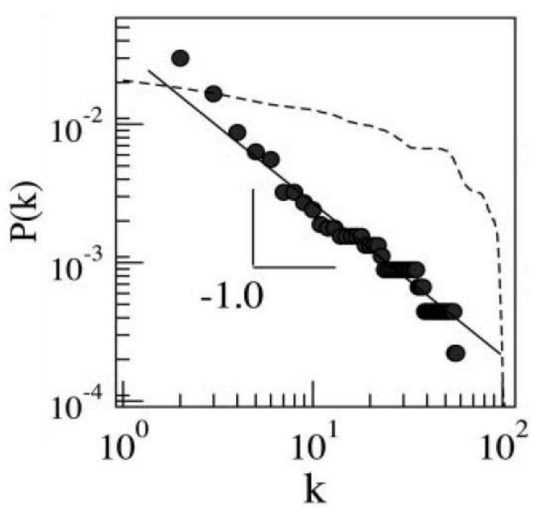
3. The Model

Healthy Speakers

- Model developed by Ferrer i Cancho and Solé (2003)
  - This model consists of a binary matrix in which columns represent words and rows represent basic ingredients for word meaning called objects of reference. A 1 in a cell means a word and an object are connected.

		Phonological Form								
		Word 1	Word 2	Word 3	Word 4	Word 5	Word 6	Word 7	Word ...	Word n
Logical Form	Obj. a	1	1	0	0	0	0	1	0	0
	Obj. b	1	0	0	1	0	0	0	0	1
	Obj. c	1	0	1	0	0	0	0	1	0
	Obj. d	0	1	0	1	0	0	0	0	0
	Obj. e	1	0	1	0	0	0	0	0	0
	Obj. f	1	1	0	0	1	0	0	0	0
	Obj. g	1	0	0	0	1	0	0	0	0
	Obj. ...	0	0	1	0	0	0	0	0	0
	Obj. m	1	1	0	0	0	1	0	0	0

- Based on this matrix, speaker and hearer entropy were calculated.
- **Entropy was lowest when the distribution of connections follows Zipf's law**: a few words have many connections and many words have few connections, according to a power law (see figure).



Number of connections *P(k)* versus rank *k* on a doubly logarithmic scale.

Aphasic Speakers
- Reduced processable entropy.
- **Entropy of the lexicon is reduced if larger differences are created between the numbers of connections of different signals.**

4. Calculations

Comparison of two modifications:
- Unavailability of all words with **one connection** (reference value)
- Unavailability of words with an **intermediate number of connections**

Outcome measure: Speaker entropy of the remaining part of the lexicon.

Words with an intermediate number of connections were removed until the entropy of the remaining lexicon was lower than the remaining part of the lexicon when all words with one connection were removed.

Speaker entropy: $H_n(S) = -\sum_{i=1}^{n} p(s_i)\log_2 p(s_i)$
where $p(s_i)$ is the probability of a signal based on its number of connections and its number of synonyms.

Results of removal of lexical classes (classes of words with the same number of connections) from the lexicon and the entropy of the remaining lexicon.

Removed frequency class From	To	Size unaffected lexicon	Entropy
none removed		400	6.244
1		**200**	**5.696**
2		333	5.929
2	3	300	5.716
2	4	280	5.390
3	5	333	5.726
3	6	324	5.616
4	7	350	5.747
4	8	344	5.666
5	10	356	5.710
5	11	353	5.658
6	14	360	5.707
6	15	359	5.685
7	17	365	5.726
7	18	364	5.695
8	22	367	5.699
8	23	366	5.673
9	27	370	5.718
9	28	369	5.691
10	32	372	5.727
10	33	371	5.694
11	36	374	5.730
11	40	375	5.692
12	44	375	5.701
12	50	374	5.665
13	50	377	5.714
13	57	376	5.681
14	66	377	5.702
14	**80**	**376**	**5.677**
15	133	376	5.723
15	200	375	5.844

Results:
- **Entropy is lowest when all words with 14 to 80 lexical connections are unavailable for lexical selection**;
- A gap of this size renders 376 out of 400 words available for lexical selection, while entropy is lower then when the 200 words with 1 connection are unavailable.

5. Proposed Experiment

Numbers of connections cannot be counted directly, so it has to be derived.
Proposed measure: **age of acquisition**.

- Words with an early age of acquisition are likely to be highly connected, while words with a late age of acquisition are only sparsely connected.
- The number of connections is likely to follow Zipf's law. If so: the number of connections is logarithmically related to the age of acquisition.

Entropy is reduced if words with an intermediate number of connections were unavailable for lexical retrieval. This leads to the following hypothesis:

Hypothesis: spontaneous speech of aphasic speakers shows fewer words with an intermediate age of acquisition compared to the spontaneous speech of healthy speakers.

6. Conclusions

- **Entropy of a lexical model is efficiently reduced if words with an intermediate number of connections are unavailable;**
- **This method is more efficient than removal of words with one lexical connection.**

Therefore, aphasic patients are hypothesized not to use words with an intermediate number of connections. It is proposed to test this hypothesis through age of acquisition.

If these hypotheses are true, then this theoretical model is capable of explaining how aphasic word finding difficulties arise, and why they occur for certain words. This method provides a quantitative way to classify aphasic patients based on their word finding difficulties.