

# The effects of frequency and morphological complexity on aphasic speech production

In this presentation, we use a case-study of aphasic speech production to compare and evaluate two different approaches to morphological competence: Pinker's dual-route model (Pinker 1991, Prasada & Pinker 1993) and item-based network models, such as Rumelhart and McClelland's connectionist framework (Rumelhart & McClelland 1986). The main difference between these two strands of research lies in their treatment of regular morphology. Pinker (1991) posits a model which combines insights from generative linguistics and connectionist approaches by claiming that regular morphological processes are represented by explicit symbolic rules, while irregular morphological processes emerge from associative networks. Adherents of connectionist models, on the other hand, claim that the available data do not warrant such a sharp distinction between regular and irregular processes, which are both represented by associative networks (e.g. Bybee 2001). Consequently, the two models make different predictions regarding the role of frequency and morphological complexity in regular morphology: the dual-route model predicts that regular morphology will only be affected by morphological complexity (i.e. complex forms will be more difficult to produce), whereas the network model predicts that it will also be affected by frequency (i.e. complex and/or infrequent forms will be more difficult to produce). Such effects may be directly observable in sentence repetition tests with aphasic patients, where forms that are difficult to produce are less likely to be repeated faithfully. Therefore, we conducted a sentence repetition test in Hungarian with an agrammatic aphasic, focussing on verbal morphology. The results suggest that frequency plays a significant role in regular morphology, which, along with a number of general arguments to be presented below, supports a model where regular and irregular processes are both represented by associative networks.

Before presenting the experiment itself, we take a brief look at some of the arguments which serve as the basis of Pinker's dual route model. First of all, Pinker claims that while frequency affects the production and perception of irregular past tense forms in English, it does not affect that of regulars. According to an experiment described in Pinker (1991), subjects produce high-frequency irregular past tense verbs faster than low-frequency ones after seeing the stem, as opposed to regular past tense verbs, where no such effect is seen. Although this seems to be a strong argument for Pinker's model, some recent experiments have produced conflicting results: Bybee (2001) reports an experiment where frequency affected the perception of both regular and irregular forms.

Pinker points out that verbs with an irregular past tense form tend to be organised into similarity clusters based on phonological features, as opposed to regular verbs, which do not form such groups. Thus, English irregular verbs ending in [ɪŋk] and [ɪŋ] usually form their past tense by ablauting the stem vowel (e.g. *ring*~*rang*), as opposed to regulars, which all take the suffix *-ed* regardless of their phonological form. Moreover, Pinker claims that Rumelhart and McClelland's similarity-based model cannot supply past tense forms for nonce verbs which have an ending not found in any existing English verb. However, this problem might be solved by extending the notion of similarity to include morphological features such as [VERB]. Novel forms which share only morphological features with other English verbs (e.g. *ploamph* or *rhumba*) would be assigned the 'default' ending *-ed*, which is not sensitive to phonological features, as opposed to novel forms which share phonological features with members of a similarity cluster (e.g. *spling*), which would either be assigned the default ending, or form their past tense like other members of the similarity cluster (e.g. *spling* → *splang*). It should also be remarked that while the above examples show a clear split between regular and irregular morphology in English, it is questionable whether such a distinction can be maintained in the description of rich morphology languages, such as Hungarian (Rebrus et al. 1996). Having seen a number of theoretical arguments for and against Pinker's dual-route model, we may now move on to describe our experiment.

We conducted a 48-item sentence-repetition test with a Hungarian agrammatic aphasic showing difficulties with lexical access. Each sentence contained a verb of medium lemma-frequency—frequencies were taken from a frequency dictionary based on a 500 million word Hungarian corpus (Halácsy et al. 2004). As our primary goal was to measure the effects of morphological complexity and frequency, the 48 verbs fall into four groups of twelve depending on whether they are morphologically simple or complex (i.e. base forms or suffixed forms), and whether they are high-frequency or low-frequency forms (by frequency, we mean the relative frequency of a form within its paradigm—high-frequency forms were usually the most frequent members of their paradigms, whereas low-frequency forms were less frequent within their paradigms). Verbs in the first group are morphologically simple high-frequency forms. Verbs in the second group are morphologically simple low-frequency forms. The third group contains forms that are morphologically complex and high-frequency. Finally, verbs in the fourth group are morphologically complex low-frequency forms. Pinker’s model predicts that base-forms will be repeated more accurately than suffixed forms, regardless of frequency, whereas item-based network models predict that the subject will be most successful at repeating high-frequency base-forms, slightly less successful at repeating high-frequency suffixed forms and low-frequency base-forms, and even less successful at repeating low-frequency suffixed forms.

The sentences obtained from the subject were classified into four groups depending on how accurately the verb was repeated: (i) accurate repetition of the verb; (ii) changes in the suffix (e.g. *ül-ünk* ‘we sit’ → *ül-nek* ‘they sit’); (iii) changes in the stem (e.g. *halad-tak* ‘they progressed’ → *alud-tak* ‘they slept’, *szervez* ‘he organises’ → *csinál* ‘does’); (iv) no repetition or unintelligible verb. The table below shows the results of the test (B-FORM stands for base-form and S-FORM for suffixed form):

	(I) ACC	(II) SUFF CH	(III) STEM CH	(IV) UNINT
H-FREQ B-FORM	8.5%	0%	10.5%	6%
L-FREQ B-FORM	6%	2%	8.5%	8.5%
H-FREQ S-FORM	6%	6%	8.5%	4.5%
L-FREQ S-FORM	0%	8.5%	10.5%	6%
SUM	20.5%	16.5%	38%	25%

The results in class (i) show the effects of both frequency and morphological complexity: high-frequency base-forms were the most likely to be repeated accurately, followed by low-frequency base-forms and high-frequency suffixed forms, with no accurate repetitions for low-frequency suffixed forms. Class (ii) shows similar tendencies: the suffix was often changed in low-frequency suffixed-forms, less often in low-frequency base-forms and high-frequency suffixed forms, and never in high-frequency base-forms. Unfortunately, there were no such tendencies in classes (iii) and (iv), which suggests that neither frequency, nor morphological complexity were relevant factors here. Thus, the results support a model where regular morphological processes are affected by both frequency and morphological complexity, that is to say, item-based network models.

Some further interesting tendencies are revealed within class (ii), if we take a closer look at the suffix changes themselves. In five cases out of six, the verb form in the subject’s answer was more frequent than the original verb form (two verb forms had to be disregarded for semantic reasons), and there was only one case where the verb form in the answer was a base-form. This means that such suffix changes are sensitive to frequency, but do not produce morphologically simpler forms, which strongly argues for item-based network models.