UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Does Frequency Determine the Storage of Compounds? Evidence from Chinese

Permalink

https://escholarship.org/uc/item/6x82k4rm

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 18(0)

Authors

Clausner, Timothy C.
Rapp, Brenda
Su, Yi-Ching

Publication Date

1996

Peer reviewed

Does Frequency Determine the Storage of Compounds?
Evidence from Chinese

Timothy C. Clausner
University of Southern California
Hedco Neurosciences Building
Los Angeles, CA 90089-2520
timc@gizmo.usc.edu

Brenda Rapp and **Yi-Ching Su**
Johns Hopkins University
Krieger Hall / 3400 N. Charles Street
Baltimore, Maryland 21218-2685
brenda@mail.cog.jhu.edu
ching@mail.cog.jhu.edu

Abstract
We investigate the role that frequency plays in the storage of morpheme pairs as lexical units. That is, do frequency differences distinguish compounds and phrases? Two hypotheses are considered: 1. Frequency of Co-occurrence (FOC): assumes that a morpheme pair will be stored if the constituents co-occur with sufficient frequency. 2. Idiomaticity (ID): assumes that a morpheme pair will be stored if the combined meaning is sufficiently unpredictable from the meaning of the constituent morphemes. We contrast two sets of morpheme pairs: *Transparent* pairs (i.e. phrases) whose meaning is generally derived from the constituents, e.g. "black coat", and *Idiomatic* pairs (i.e. compounds) whose meaning is not entirely derived from the constituents, e.g. "blackboard". If the two sets are matched for frequency of morpheme co-occurrence <u>and</u> for the frequency of the constituent morphemes, then FOC predicts that the two sets should be comparable on relevant behavioral measures since frequency is held constant. In contrast, ID predicts behavioral consequences based on the differences in idiomaticity even for frequency matched sets.

These hypotheses are not readily tested in English which marks compounds by both phonological stress and the number of orthographic "words". However, Mandarin Chinese does not use phonology or orthography to distinguish between idiomatic and transparent morpheme pairs. The compound "bai tsai" (lit., white vegetable) does not mean any white vegetable, but has the slightly idiomatic meaning, a specific kind of light green cabbage. The phrase "bai qi" (lit., white paint) is transparent in meaning, referring to any paint that is white.

Method. Twenty native speakers of Mandarin Chinese saw compounds, phrases, and nonsense morpheme pairs and were asked to decide whether each pair was meaningful, responding with a key press as quickly and accurately as possible. All pairs were matched for constituent morpheme frequency. Compounds and phrases were also matched for frequency of morpheme co-occurrence.

Results

Analysis 1. The FOC hypothesis predicts that morpheme pairs matched for frequency of co-occurrence should produce similar RTs and error rates, regardless of semantic factors such as idiomaticity. The ID hypothesis predicts that idiomatic morpheme pairs will be stored and transparent pairs will not. Storage should provide an RT and accuracy advantage to lightly idiomatic vs. transparent morpheme pairs. However, results reveal that responses were significantly faster and more accurate for lightly idiomatic vs. transparent pairs. Therefore, FOC is not supported.

Analysis 2. A weaker version of FOC predicts that idiomaticity plays some role but frequency of co-occurrence is nonetheless a sufficient condition for storage of a combined meaning—i.e. high frequency morpheme pairs would be stored regardless of idiomaticity. This predicts no RT or error rate differences between matched high-frequency compounds and phrases. Again, however, RT and error rate differences for the two sets were significant. This indicates that frequency of co-occurrence is not a sufficient condition to establish storage of morpheme pair meaning.

Analysis 3. In order to examine if idiomaticity alone is sufficient for storage we compared results for matched low-frequency co-occurring morpheme pairs. We observed that for low-frequency pairs responses were more accurate for compounds than phrases, although RTs did not differ significantly. Possible explanations for the lack of an RT effect are: (i) idiomaticity and frequency of occurrence may both determine the storage of a compound—i.e. idiomaticity may be a necessary but not a sufficient condition for storage, (ii) we chose problematic low frequency pairs. In fact, the low frequency Verb-Noun pairs differ in RTs and error rates from Adjective-Noun pairs.

Conclusions
From current results we conclude that the frequency of morpheme co-occurence is insufficient for determining a storage difference between idiomatic and transparent morpheme pairs. Idiomaticity is apparently necessary for storage although there is some indication that a certain frequency of co-occurrence may also be required. However, because of possible confounds involving the grammatical category of the stimuli, future work is required to determine if frequency plays any significant role in the lexicalization/storage of morpheme pairs.