Studying Morphological Processing via English Lexical Decision Tasks

Gregory Theos, BA&Sc Linguistics & Computer Science | Advisor: Dr. Timothy O'Donnell

Department of Linguistics, McGill University

Research Questions

- Is it possible to tell if redundant storage of inflectional processes occurs in processing of words?
- If it is possible to tell, does redundant storage actually occur?

Background

In morphology, the study of words and word-building, a word can either be **morphologically simple** or **morphologically complex**. Morphologically complex words are made up of multiple parts which each have a meaning on their own. Complexity occurs in compounds such as in "un-true", and also results from grammatical processes such as verb conjugation, generally called **inflection**. Inflection can be seen in "blow-s", where the affix -s attaches to the verb *blow*, signaling that the verb is in its 3rd person singular form. We can describe any of these processes of inflection as **regular** or **irregular**, contrasting ones that arise from a predictably regular rule (you can take practically any English verb and add -s to it) with ones that arise from no clearly predictable rule (such as "I am" becoming "he is").

With the terminology settled, we can posit that, in the mind, any particularly inflectional process can be **computed** or **stored**. A computed process will entail the inflection being assembled "live", as it were, akin to multiplying to numbers by getting out a pencil and paper without knowing in advance what the answer might be. In contrast, the process could also have been stored, and its result accessed whole from the statement of the problem, akin to memorizing multiplication tables. Therefore we have the possibilities in figure 1.
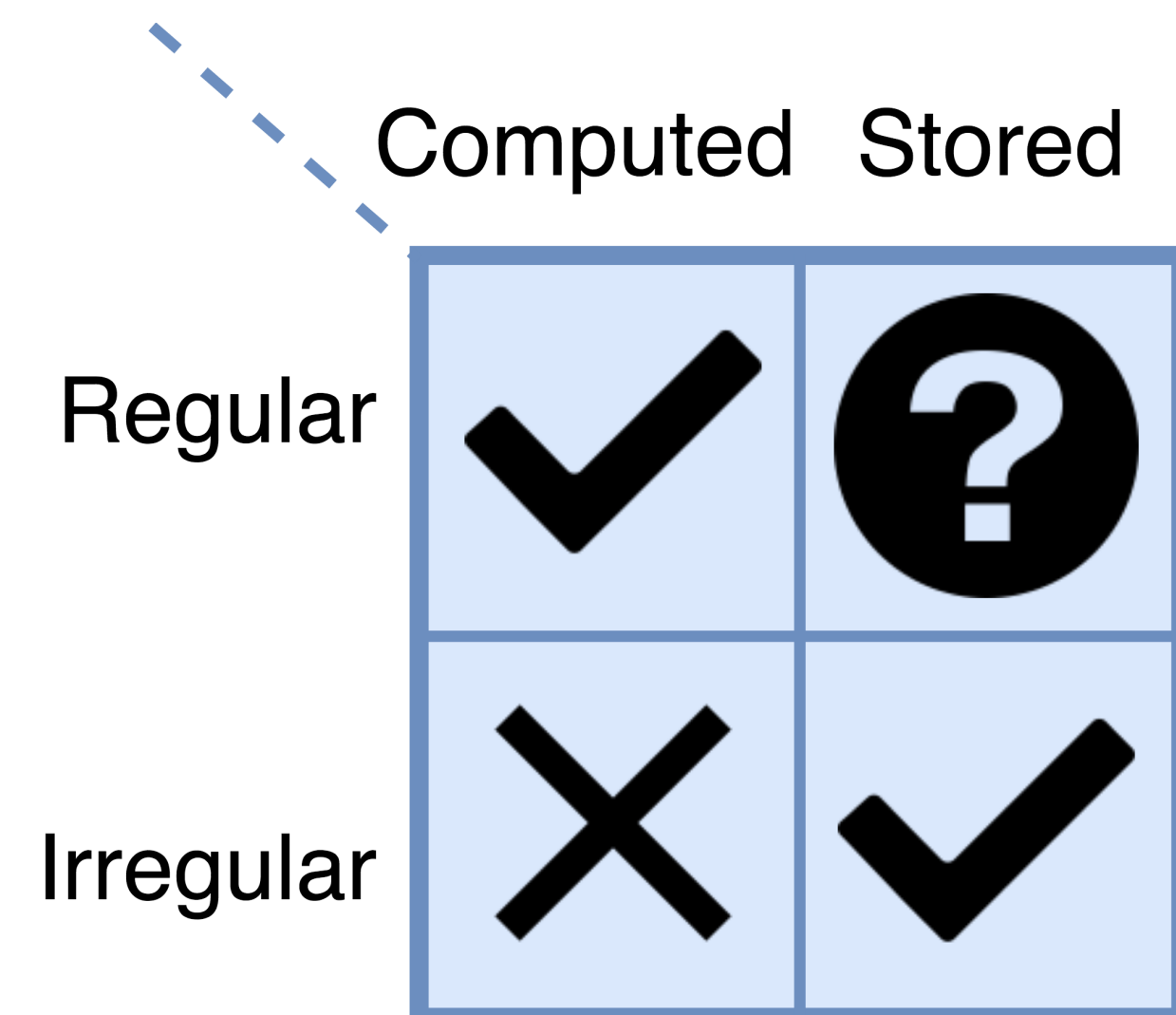

Figure 1: possibilities for computed or stored inflectional processes

Background

Evidence for/against the cells above is quickly evaluated here:

- Computed and Regular – This type of process must be possible, because humans can and do apply regular inflectional processes to words they have never seen before.
- Computed and Irregular – This type of process is impossible, since by its very definition an irregular rule cannot be predicted, and therefore cannot be computed on the fly.
- Stored and Irregular – This type of process must happen, since irregular forms exist, and as we have established above, cannot be computed.
- Stored and Regular – There is no *a priori* way to tell whether this type of process takes place. Although it is redundant, there is nothing stopping it from taking place, and we can easily imagine that it might save the brain time in certain situations.

Methods & Materials

We can use evidence from lexical decision tasks, in which subjects are presented with a stimulus (usually a written word) on a screen and asked to judge whether the stimulus is a real word or not. Their judgements and reaction times are recorded. In particular, a large publicly available corpus of these experiments, the English Lexicon Project, was used.

The general logic of the modeling setup was the following: It is well established that word frequencies have a heavy influence on reaction times in lexical decision tasks. Particularly, the frequencies that matter are the frequencies of what is being accessed from memory. Therefore, if the word is being computed, there will be two accesses to memory of the two morphemes in the word; if the word is being stored, there will be only one access, namely to the memory trace of the whole word. This is illustrated in figure 2 below.

Therefore a statistical model can be fit to reaction time data, with predictors for whole-form frequency as well as for base and affix frequency.
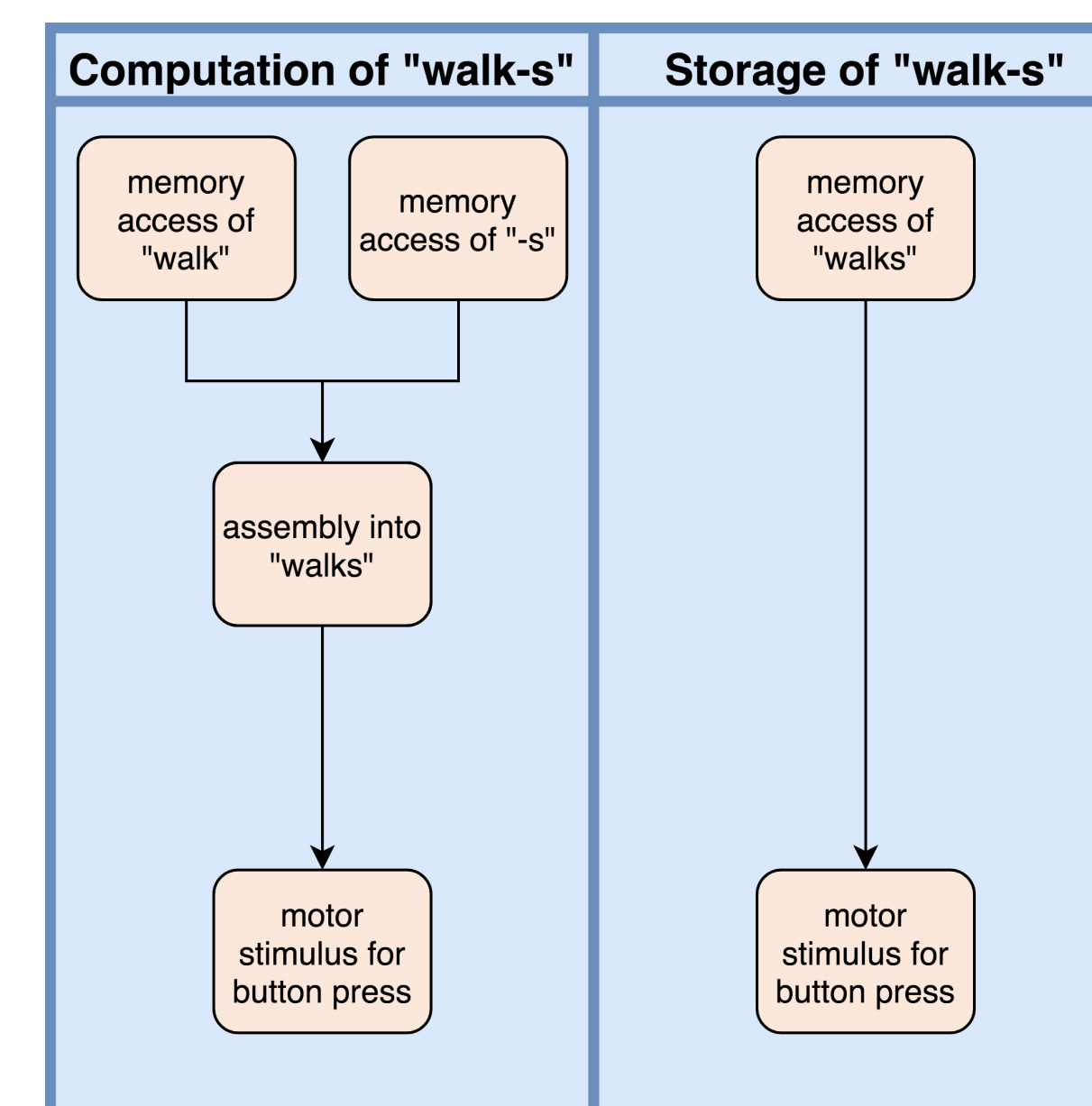

Figure 2: flowchart depicting roughly the process of perception and reaction to a stimulus "walks" under two different scenarios

Further Work

A baseline model has already been constructed, attempting to replicate the result of a related paper by Constantine Lignos and Kyle Gorman from 2012. Additional modeling will be done in the fall 2018 semester with the following potential innovations:

- Use Bayesian modeling to assign a likelihood to whether a word is computed or stored
- Use Monte Carlo sampling to attempt to establish a correct threshold for storage vs. computation

References

[1] D. Balota and M. et al Yap. The english lexicon project. *Behavior Research Methods*, 2006.

[2] C Lignos and K Gorman. Revisiting frequency and storage in morphological processing. In *Proceedings of annual meeting of the Chicago Linguistic Society*, pages 447–461. Chicago Linguistic Society, January 2012.

Acknowledgements

Contact Information

- Email: gregory.theos@mail.mcgill.ca