



HAL
open science

An algorithm for analyzing cloze test results

Jonathan Mirault, Stéphanie Massol, Jonathan Grainger

► **To cite this version:**

Jonathan Mirault, Stéphanie Massol, Jonathan Grainger. An algorithm for analyzing cloze test results. *Methods in Psychology*, 2021, 5, 10.1016/j.metip.2021.100064 . hal-03940613

HAL Id: hal-03940613

<https://hal.science/hal-03940613v1>

Submitted on 16 Jan 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/353124562>

An algorithm for analyzing cloze test results

Article in *Methods in Psychology* · July 2021

DOI: 10.1016/j.metip.2021.100064

CITATION

1

READS

134

3 authors:



Jonathan Mirault

Aix-Marseille Université

31 PUBLICATIONS 114 CITATIONS

[SEE PROFILE](#)



Stéphanie Massol

Université Lumière Lyon 2

22 PUBLICATIONS 236 CITATIONS

[SEE PROFILE](#)



Jonathan Grainger

Aix-Marseille Université

425 PUBLICATIONS 22,350 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Letter perception [View project](#)



Spatial Attention and Reading [View project](#)



An algorithm for analyzing cloze test results

Jonathan Mirault^{a,*}, Stéphanie Massol^b, Jonathan Grainger^a

^a Laboratoire de Psychologie Cognitive (UMR 7290), Aix-Marseille Université & Centre National de la Recherche Scientifique, Marseille, France

^b Laboratoire d'Étude des Mécanismes Cognitifs, Université Lumière Lyon 2, Bron, France

ARTICLE INFO

Keywords:

Cloze test
Sentence comprehension
Word identification

ABSTRACT

We present an algorithm for the automatic processing of cloze test results that are traditionally analyzed and adjusted by hand. The algorithm accurately includes inflected and derived forms of the target word as correct responses, as well as minor spelling mistakes and typographical errors. We provide a test of the algorithm against the coding of one expert human observer and find that the algorithm performs with equal accuracy. We further examine the utility of the algorithm for the analysis of data involving word identification with typed responses. The routines are implemented in R-software and are freely available.

1. Introduction

The cloze test is a technique used in psycholinguistic studies of sentence comprehension in order to evaluate the predictability of words given a preceding context.¹ Thus, for example, in a study investigating how predictability impacts on reading behavior, the processing of more predictable words can be compared with less predictable words preceded by an identical context (e.g., The hunter was afraid of the _; lion vs. deer). In order to evaluate the predictability of words in a given context, an independent group of participants are first tested with the incomplete sentence contexts (e.g., The hunter was afraid of the _) and asked to complete the sentence frame with a single word (i.e., the first word that comes to mind). The cloze probability of a given word (typically a word to be tested in a subsequent experiment, such as “lion” and “deer” in the example) is then calculated as the number of times the word is given in response divided by the number of possible responses (i.e., the number of participants).

It is important to note that in recent years the number of studies investigating the impact of word predictability on word identification in a sentence context has increased (e.g., Brothers and Kuperberg, 2021; Chang et al., 2020; Cutter et al., 2020; Rommers et al., 2020; Rommers and Federmeier, 2018; Sereno et al., 2020). This points to a growing need for a formalized means of analyzing the results obtained in cloze tests in order to provide a precise and standardized measure of predictability. Indeed, current practice varies from one study to another notably in terms of whether or not derived or inflected forms are

considered as correct, or whether or not obvious spelling mistakes and typographical errors are taken into consideration. Moreover, the number of participants performing the cloze test varies considerably from one study to another (e.g., from 20 to 100 participants). Clearly, the greater the number of participants the more reliable the results will be. Thus, given the relative ease with which participants can be recruited on-line, any limitation in sample size might well be driven more by consideration of the amount of time required to analyze the data. Hence the present contribution.

To our knowledge, analysis of cloze test results is typically done by hand. This is problematic for two reasons. First, it can be a very tedious process, particularly given that a large number of participants is desirable in order to obtain reliable cloze test results. Second, hand coding is of course prone to human error and subjective influences. In the present commentary we describe an algorithm for the automatic processing of cloze test results, and we compare the output of the algorithm with that of one skilled human observer. We apply the algorithm to French, but an extension to other languages such as English can be straightforwardly implemented simply by changing the reference lexicon and the list of affixes that are used by the algorithm.

2. The algorithm

The algorithm was developed in the R programming environment (R Core Team, 2018), is available on the Open Science Framework (OSF) at osf.io/9rhyg (c.f. “Algo_1.R” and “Algo_2.R” in the OSF package) and is

* Corresponding author.

E-mail address: Jonathan.mirault@univ-amu.fr (J. Mirault).

¹ We follow tradition here by referring only to the impact of preceding context, but we note that parallel versions of the cloze test can be straightforwardly implemented and might be more appropriate in certain circumstances (e.g., Snell and Grainger, 2019).

represented in the form of a flow-chart in Fig. 1 below. Prior to applying the algorithm all accents were removed from responses containing accented letters (i.e., “élève” became “eleve”). In order to handle compound words, we included hyphens and spaces as characters with the same status as letters. In the following description of the algorithm the term “response” refers to participants’ answers in the cloze test, and the word “target” refers to the experimenter’s expected response – that is, the completion of the sentence frame to be used in a planned experiment. The algorithm involves the following steps:

1. Is the response a word in the reference lexicon? For the French version of the algorithm the reference lexicon was the Lexique3 database (New et al., 2005).
2. If the **response is a word**, then it is compared with the target, and if there is an exact match with the target then the response is accepted.
3. If the **response is a word** that does not provide an exact match with the target, then all affixes are removed from the target and response using a list of prefixes and suffixes of the target language. For the French version, we use the affixes listed in MorphoLex FR (Mailhot et al., 2020) while retaining Lexique3 as the reference lexicon. Then the algorithm checks again if the two strings match. This allows responses such as “reading” or “re-read” to be accepted for a target such as “reads”.²
4. If the **response is not a word**, then the orthographic distance with the target is calculated. Orthographic distance was measured using the *stringdist* (Van der Loo, 2014) package in R. This measures the Optimal String Alignment (OSA), which is a Levenshtein distance that includes transpositions of adjacent characters as one count. Thus, single letter insertions, deletions, substitutions, and adjacent letter transpositions were each considered as a 1 distance value and were combined to provide the total distance between the response and the target. We decided to include transpositions since these are quite common typographical errors. If the orthographic distance is less than 3 then the response is accepted, otherwise it is rejected.

3. Output of the algorithm

For a given target, each time a response is accepted a counter is incremented. The cloze probability associated with a given target corresponds to the number of accepted responses determined by the algorithm, divided by the number of participants (in this way, an absence of response is included as a response).³

4. Comparison of the algorithm with a human observer

The dataset for this comparison was the cloze test results reported in Massol et al. (2021). This test involved a total of 480 French 4-word sentences. On each trial, the first 3 words of the sentence (e.g., *votre film est __/your film is __*) were presented to participants and they were asked to type the first word that came to their mind as a likely continuation of the sentence. Participants were informed that a response was required on each trial. The cloze test was performed online using Google Forms. Before calculating the cloze probability of each of the 480 target words (i.e., the words that completed each sentence to be tested in a subsequent experiment), spelling mistakes and typographical errors were corrected by hand by the human observer. In addition, we applied the algorithm to the same set of 480 sentences and compared the output of the algorithm with that of the human observer. We report the

² We note, nevertheless, that in the absence of a morphologically tagged reference lexicon the algorithm will also accept “reads” as a correct response for “ready”. However, we doubt that many participants will respond “reads” instead of “ready” in a cloze test.

³ Alternatively, the total number of responses (excluding no responses) could be used as the denominator.

proportion of trials where the algorithm and the human observer diverged in terms of the calculated cloze probability. We found that on 97.3% of trials the difference was less than 0.1% (mean Human = 0.035, SD = 0.081; mean Algorithm = 0.036, SD = 0.082; $t = 0.41$).

5. Application of the algorithm for measuring accuracy in word identification

We also provide a version of the algorithm where there is no expected response (c.f. “Algo_3.R” in the OSF package). This version outputs the probability that a given response occurred among all responses that were recorded. This unconstrained version of the algorithm applies the same principles as the main algorithm but where any word, or closely matching nonword is accepted. It can be applied to provide more flexible encoding of response accuracy compared with an exact match in any experiment involving word identification.

Indeed, many studies of word recognition, or word-in-sequence identification require a response typed on a keyboard. These data are then automatically analyzed using a simple match-to-target algorithm (c.f. “Algo_4.R” in the OSF package), with a perfect match counting for one, and otherwise zero.⁴ In certain experiments it might be desirable to include typographical errors and spelling mistakes as correct responses, and also possibly inflected and derived variants of the target word. The unconstrained version of our algorithm, described above, does precisely that. Here we provide an example of the number of data points that are recovered by applying the algorithm compared with an exact match procedure. To do so, we compared 3 datasets (obtained with 3 different groups of participants) from a word-in-sequence identification experiment with 480 items (Massol et al., 2021). The percentage of trials where the algorithm recovered additional data points relative to the exact match procedure was: 65.83% for the first dataset, 76.45% for the second dataset and 57.79% for the third dataset.

6. Conclusions and future directions

We describe an algorithm that can be used to automatically process the results of a cloze test. This obviously saves a large amount of time compared with hand coding responses and has the added advantage of avoiding potential human errors. The algorithm accepts minor spelling mistakes and typographical errors as correct responses, as well as inflected or derivational variants of the expected response (e.g., “reads” instead of “reading”). The output of the algorithm did not differ significantly from the hand-coding performed by one expert human observer.

An equally useful application of the algorithm is for coding typed responses in a word identification experiment in order to accept as a correct response typing errors and affixed forms when this is desirable. Such automatic correction of typed responses is particularly useful for on-line experiments given the much larger participant samples that are typically tested in order to compensate for greater noise compared with laboratory experiments.

Finally, we note, following the recommendation of a reviewer, that the algorithm could be used in combination with a human observer by coding for the different types of non-exact matches detected by the algorithm (unacceptable error, acceptable spelling error, acceptable morphological variant). The human observer would then just have to check these specific cases as opposed to having to check all responses. We have implemented this possibility and provided it as an option in the two first algorithms (“Algo_1.R” and “Algo_2.R”). The same reviewer also pointed to the future possibility of training the algorithm using hand-coded human data in order to optimize its performance.

⁴ Note also that a simple exact match procedure will consider “house”, “House”, and “HOUSE” as different strings.

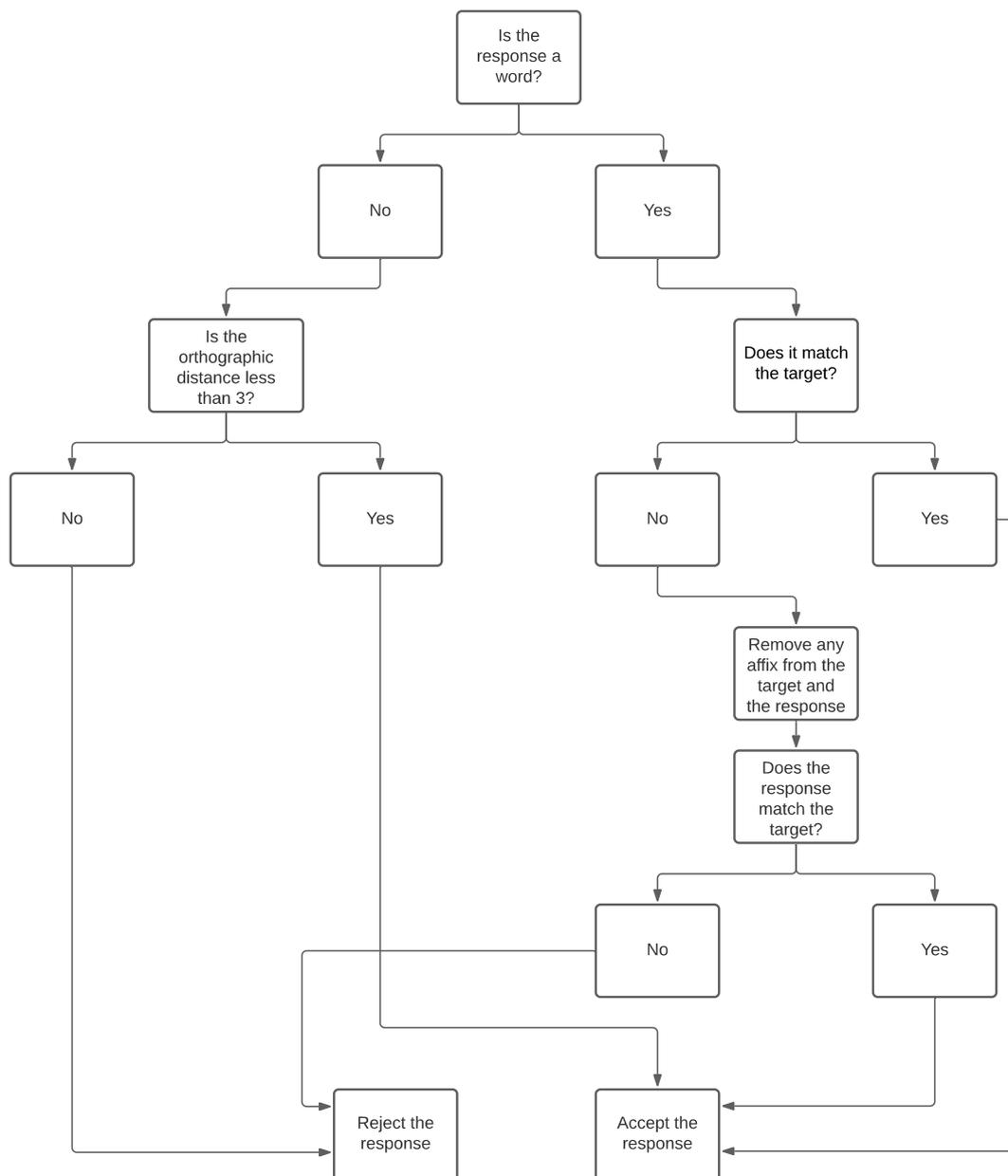


Fig. 1. Flow-chart of the algorithm for analyzing responses in the cloze test.

Disclosure statement

No potential conflict of interest was reported by the Authors.

Funding

This work was supported by European Research Council (742141).

Author’s contributions

Jonathan Mirault: Design, Script, Writing. Stéphanie Massol: Hand-coding, Writing. Jonathan Grainger: Design, Writing.

Acknowledgments

This work was funded by the European Research Council (grant 742141) awarded to Jonathan Grainger. We thank Marc Brysbaert for his excellent advice concerning possible extensions of the algorithm.

References

Brothers, T., Kuperberg, G.R., 2021. Word predictability effects are linear, not logarithmic: implications for probabilistic models of sentence comprehension. *J. Mem. Lang.* 116, 104–174.

Cutter, M.G., Martin, A.E., Sturt, P., 2020. The activation of contextually predictable words in syntactically illegal positions. *Q. J. Exp. Psychol.* 73, 1423–1430.

Chang, M., Zhang, K., Hao, L., Zhao, S., McGowan, V.A., Warrington, K.L., Paterson, K.B., Wang, J., Gunn, S.C., 2020. Word predictability depends on parafoveal preview validity in Chinese reading. *Vis. Cognit.* 28, 33–40.

Mailhot, H., Wilson, M.A., Macoir, J., Deacon, S.H., Sánchez-Gutiérrez, C., 2020. MorphoLex-FR: a derivational morphological database for 38,840 French words. *Behav. Res. Methods* 52, 1008–1025.

Massol, S., Mirault, J., Grainger, J., 2021. The contribution of semantics to the sentence superiority effect (submitted).

New, B., Pallier, C., Ferrand, L., 2005. *Manuel de Lexique 3*. *Behav. Res. Methods Instrum. Comput.* 36 (3), 516–524.

R Core Team, 2018. R: a Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna. <https://www.R-project.org>.

Rommers, J., Dell, G.S., Benjamin, A.S., 2020. Word predictability blurs the lines between production and comprehension: evidence from the production effect in memory. *Cognition* 198, 104206.

- Rommers, J., Federmeier, K., 2018. Predictability's aftermath: downstream consequences of word predictability as revealed by repetition effects. *Cortex* 101, 16–30.
- Sereno, S.C., Hand, C.J., Shahid, A., Mackenzie, I.G., Leuthold, H., 2020. Early EEG correlates of word frequency and contextual predictability in reading. *Language, Cognition and Neuroscience* 35, 625–640.
- Snell, J., Grainger, J., 2019. Readers are parallel processors. *Trends Cognit. Sci.* 23, 537–546.
- Van der Loo, M.P., 2014. The stringdist package for approximate string matching. *The R Journal* 6, 111–122.