Acquiring Agglutinating and Fusional Languages Can Be Similarly Difficult: Evidence from an Adaptive Tracking Study

Svenja Wagner (s1581727@sms.ed.ac.uk),
Centre for Language Evolution, The University of Edinburgh,
3 Charles Street, Edinburgh, EH8 9AD, UK
Kenny Smith (Kenny.Smith@ed.ac.uk),
Centre for Language Evolution, The University of Edinburgh,
3 Charles Street, Edinburgh, EH8 9AD, UK
Jennifer Culbertson (Jennifer.Culbertson@ed.ac.uk)
Centre for Language Evolution, The University of Edinburgh,
3 Charles Street, Edinburgh, EH8 9AD, UK

Abstract

Research on the acquisition of morphology commonly predicts that agglutinating systems should be easier to learn than fusional systems. This is argued to be due to compositional transparency: the mapping between morphemes and meanings is one-to-one in agglutinating systems, but not in fusional systems. This is supported by findings in first and second language learning (Goldschneider & DeKeyser 2001, Slobin 1973), typology (Dressler 2003, Haspelmath & Michaelis 2017), and language evolution (Brighton 2002). We present findings from a series of artificial language learning experiments which complicate this picture. First, we show that when only two features (e.g., NOUN CLASS and NUMBER) are morphologically encoded, the learnability of fusional and agglutinating systems does not differ significantly. This finding holds when learners are given an additional cue to morpheme segmentation–which in principle should make the agglutinating system easier. However, the error patterns of the two groups provide some evidence that learners might have a bias for transparent structures. Our results suggest that the advantages of agglutinating over fusional systems may be overstated, particularly when a small number of features are encoded. Since agglutinating systems likely bear additional costs (e.g., segmentation, longer word length, and the online cost of mapping between morphemes and meanings), such systems do not guarantee learning ease under all circumstances.

Keywords: language acquisition; morphology; agglutinating; fusional; artificial language learning; transparency

Introduction

Classification of languages into morphological types is a commonly used parameter in language typology. Morphological type structures vary within and between languages, and they change over time. One key distinction is between fusional and agglutinating types. The distinction between these two is based on the ratio of morphemes to meaning, where a morpheme is defined as "the smallest meaning-bearing unit of language" (Kortmann, 2005). In fusional languages, morphemes typically express more than one meaning. For example, the German verb *spielst* ('you play') has the suffix –s*t*, which together expresses present tense, second person, and singular number. In comparison, morphemes in agglutinating languages typically only carry a single meaning. For example, the Turkish verb *konuşuyorsunuz* ('you speak') has three suffixes, *-yor*, *-sun*, and *-uz* individually expressing the same pieces of information (present tense, second person, plural number).

While both morphological types are well attested among the languages of the world, it has been proposed that fusional and agglutinating systems may differ in terms of learnability. In particular, it has been claimed that the more meanings a single morpheme carries, the less transparent it is, and therefore the more difficult it is to learn (e.g., Goldschneider & DeKeyser, 2001; Don, 2017; Haspelmath & Michaelis, 2017). For the purpose of this study, we use transparency to mean one-to-one correspondence between a form and its meaning (Don, 2017). Because agglutinating morphology is by definition more transparent, agglutinating systems should be easier to acquire, while fusional systems where a single morpheme encodes multiple meanings should be more difficult.

Support for this idea comes from research on both first and second language acquisition. In first language acquisition, a number of classic studies on morpheme order of acquisition in children have implicated transparency as part of the explanation for why certain English morphemes are acquired earlier than others (Brown, 1973; de Villiers & de Villiers, 1973; Dulay & Bert, 1974). For example, *–s* as in *plays* (which expresses 3rd person, singular, and present tense) is learned later than *–ing* (progressive). These studies build on more general claims relating transparency to ease of acquisition in children (e.g., Slobin, 1973)

More recent work has extended these findings to a number of other languages. For example, Sultana, Stokes, Klee, and Fletcher (2016) argue that the level of transparency of morphological forms predicts the order of acquisition in Bengali. Hengeveld and Leufkens (2018) point out that Turkish children generally master the agglutinating morphology of their language by the age of 3, whereas Dutch children have not yet acquired the fusional verbal system of their language at that age. This is in line with Dressler (2003), who reports earlier acquisition of morphology by children in Turkish than in English.

Second language acquisition research has echoed the role of transparency in morphological learning. In a meta-analysis of 14 studies on L2 acquisition of English, Goldschneider and DeKeyser (2001) show that transparency correlates with earlier acquisition. For example, L2 learners, like children, acquire the English morpheme *–s* relatively late.

Finally, there is a clear relationship between agglutinating systems and the more general feature of compositionality. In compositional systems, complex signals are formed by combining meaning-bearing parts, with the meaning of the whole being a function of the meaning of the parts; this can be contrasted with holistic systems in which such re-combinable subparts do not exist, the relationship being between whole meanings and unanalyzable signals. A large body of research on the evolution of compositionality in language connects it to learnability (Brighton, 2002; Kirby, Cornish & Smith, 2008; Kirby, Tamariz, Cornish & Smith 2015, a.o.): compositional systems are simpler in that they have a shorter encoding length and are more compressible, making them simpler in a cognitively-relevant sense and therefore easier to learn; compositional systems also permit generalization to unseen meanings and signals. These same characteristics hold for agglutinating systems, suggesting that they too should be easier to learn.

To summarize, various lines of evidence suggest that agglutinating languages should be easier to learn than fusional languages. The inherent transparency and regularity of agglutinating forms, the higher frequency of a given morpheme in the system, and the possibility to generalize all point to a learnability advantage of these systems. However, in many cases, it is difficult to disentangle transparency from other features of the system. Most obviously, agglutinating systems often use *more* morphology overall, which could in principle also serve to obscure this advantage. However, Dressler (2003) argues that the systematic use of morphology in agglutinating languages relative to fusional ones may in fact serve to clue learners into its importance, triggering earlier learning. In this paper, we report a series of artificial language learning experiments which allows us to test the above claims by directly comparing agglutinating to fusional systems, while controlling for systematicity of morpheme use, and number of morphemes across conditions.

Experiment 1

We tested whether learners are faster at acquiring agglutinating systems compared to fusional systems by exposing participants to nouns encoding two binary features, one for NUMBER (singular/plural) and one for CLASS (animate/inanimate). Crucially, we held the number of morphemes to be learned constant across both conditions.

Methods

Participants. 80 participants were recruited on Amazon Mechanical Turk, all self-reported as English native speakers. They were paid $4 for their time. Participants were randomly assigned to one of the conditions described below (38 in the fusional and 42 in the agglutinating condition).

Materials. The language consisted of 96 nouns, referring to objects, and four suffixes, encoding NOUN CLASS (animate and inanimate) and NUMBER (singular and plural). Animate entities were always animals and inanimate entities were everyday objects such as household items and pieces of clothing. All stems were monosyllabic and adhered to English phonotactics. Morphemes used for both languages were identical: *-mu*, *-ka*, *-pi*, *-lo*. In the fusional condition, each of the four morphemes expressed one value for both features: animate+singular, animate+plural, inanimate+singular, inanimate+plural. For example, in Figure 1, *spur* is the noun stem, and *-ka* indicates animate+singular. In the agglutinating condition, the four morphemes each expressed a single value of NUMBER *or* CLASS. For example, in Figure 2, *foog* is the stem, *–ka* indicates inanimate, and *–mu* indicates plural. Note that the stem was directly followed by the CLASS morpheme, which was followed by the NUMBER morpheme. Mappings between morphemes and meanings were randomized across participants. Note that because we use the same set of morphemes in both languages, the words are longer in the agglutinating condition (by one syllable). This is a general characteristic of agglutinating languages, where words tend to be longer than in fusional languages.

Procedure. Participants were instructed that they would be learning part of a new language. On each trial (Figures 1, 2), participants saw an image and were given a choice of four words that could describe it. The four choices always represented the same stem with four possible grammatical combinations of affixes. Participants were instructed to click on the word that they thought correctly described the picture. Immediate feedback was given in every trial: the correct answer was highlighted with color, and the audio of the correct word was played aloud. The study consisted of 96 trials each of which displayed a unique picture. Therefore, no image or stem was ever repeated (and participants were not required to learn the mappings between images and stems). Each combination of grammatical meanings occurred as the correct choice 24 times in total. At the end, participants completed a short questionnaire.

Results. The design of our experiment aimed to compare performance across conditions over time. Since participants were necessarily guessing at the beginning, we expect performance in both conditions to be similar early on, but to potentially diverge over trials as they learned. Figure 3 shows mean accuracy across conditions by trial. As expected, participants generally improved over trials. However, performance appears to improve at a similar rate across conditions. Mean accuracy across all trials for the agglutinating condition was 0.65 (SD=0.24), for the fusional condition 0.60 (SD=0.24). To test whether the rate of improvement differed between conditions we fit a logistic mixed-effects regression model, predicting correct answer by condition, trial (coded 0-95), and their interaction.[1] Condition was dummy-coded, with agglutinating as the reference level. The by-item intercept was removed because the model failed to converge. The model revealed a significant effect of trial number (b=0.04, SE=0.01, p<0.001), indicating that participants improved their accuracy over the course of the experiment, but no effect of condition (b=0.14, SE=0.18, p=0.42) and most importantly, no condition by trial number interaction (b=-0.01, SE=0.01, p=0.20). The latter would have indicated a difference in the rate of learning in one or the other condition, indicating a learnability advantage.

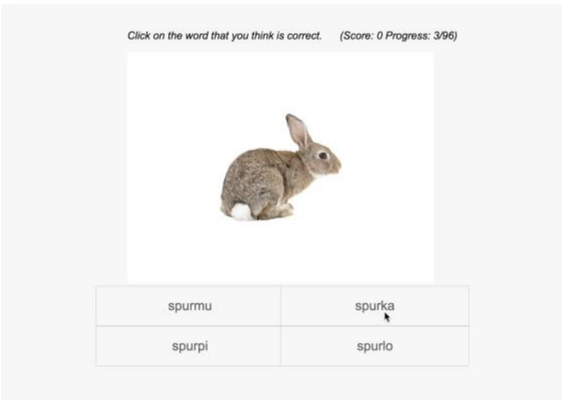

Figure 1: Example trial in Experiment 1, fusional condition. This trial shows an animate, singular object.

Figure 2: Example trial in Experiment 1, agglutinating condition. This trial shows an inanimate, plural object.

We conducted an exploratory analysis of participants' errors to investigate whether the highly similar overall performance masked a difference in error type between the two conditions. As described above, each of the four choices given in every trial constituted a different combination of grammatical meanings. Incorrect responses could either reflect the participant selecting a marker which was appropriate to the CLASS of the noun (e.g. selecting a morpheme marking animacy for an animate referent) but the wrong NUMBER (e.g. selecting a plural morpheme for a singular noun), selecting the wrong CLASS but the correct NUMBER, or selecting a morpheme which was incorrect for both CLASS and NUMBER. The rates for these three classes of error (correct CLASS only, correct NUMBER only, neither correct) are shown in Figure 4. The proportion of errors reflecting correct CLASS appears to be greater in the fusional condition. This impression is confirmed by a logistic mixed-effects regression model testing whether the proportion of CLASS correct only responses was significantly different between conditions. We ran the model predicting correct CLASS in the subset of the data with incorrect answers, including fixed effects of condition, trial number and their interaction. The by-item intercept was removed due to convergence errors. The model revealed a significant effect of trial (b=0.02, SE=0.01, p=0.003) no significant effect of condition (b=-0.02, SE=0.23, p=0.92), and a significant interaction between condition and trial (b=0.02, SE=0.01, p=0.04).

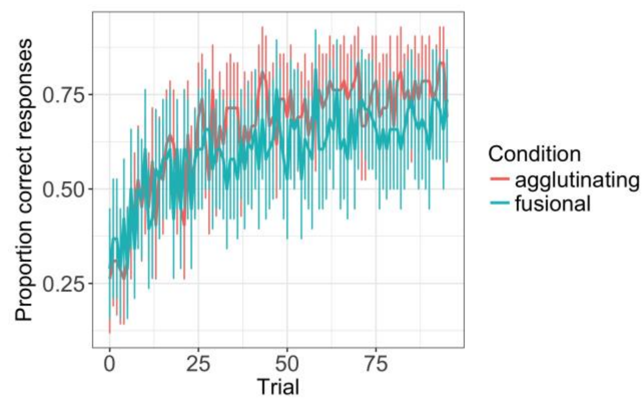

Figure 3: Mean accuracy by trial by condition in Experiment 1.

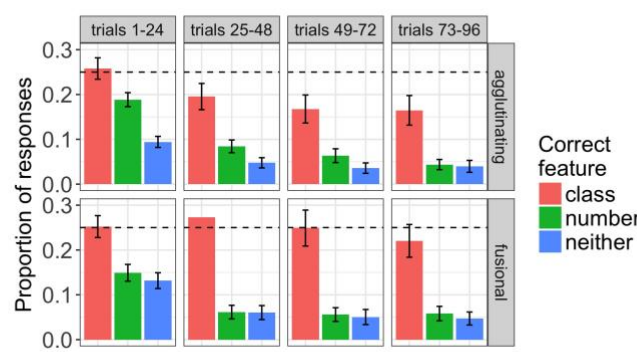

Figure 4: Classification of incorrect choices by correct feature (CLASS correct NUMBER wrong, NUMBER correct CLASS wrong, or neither correct) by trial block by condition

[1] All models were run using the package lme4 in R (Bates 2010). Unless otherwise noted, models included random by-participant and by-item (picture) intercepts, and by-participants slopes for the effect of trial.

in Experiment 1. Trial number is binned for readability. Note that the y-axis range does not display the full range.

This suggests that participants in the fusional condition and agglutinating condition diverged over time in their tendency to choose an answer in which only the CLASS morpheme was correct–while such errors decline in the agglutinating condition, they remain at a fairly constant level in the fusional condition. One possibility is that this reflects a bias for transparency: participants in the fusional condition may have been searching for a single feature with four values, rather than two binary features. While NUMBER is unambiguously binary (one vs. two), the stimuli could in principle encode more fine-grained distinctions of CLASS. In the post-test questionnaire, some participants indeed reported such a strategy, for instance, distinguishing land vs. sea animals and household items vs. clothing. They subsequently tried to map each of these four CLASS values onto one morpheme, ignoring NUMBER altogether. However, this analysis was performed post-hoc since the given distribution of answer types was unexpected. We therefore replicated the experiment.

Experiment 2

Methods

Participants. 100 participants were recruited on Amazon Mechanical Turk, all self-reported as English native speakers. They were paid $4 for their time. Participants were randomly assigned to one of the two conditions (48 in the fusional and 52 in the agglutinating condition).

Materials. Stimuli were identical to those of the previous experiment.

Procedure. The procedure was identical to Experiment 1.

Results. Figure 5 shows mean accuracy by trial across conditions. As in Experiment 1, participants generally improved from the start to the end as expected, and overall performance appears to be similar across conditions (agglutinating M=0.64, SD=0.26; fusional M=0.68, SD=0.25). We ran a model predicting correct answer by condition, trial (coded 0-95), and their interaction. Condition was dummy-coded, with agglutinating as the reference level. The model revealed a significant effect of trial number (b=0.04, SE=0.01, p<0.001), indicating that participants improved their accuracy over the course of the experiment, but no effect of condition (b=-0.10, SE=0.17, p=0.57) and no condition by trial number interaction (b=0.01, SE=0.01, p=0.25). Again, the latter would have indicated a more rapid improvement in one or the other condition, and thus a learnability advantage.

We repeated our analysis of error types across conditions (Figure 6). In this case, the model revealed a significant effect of trial (b=0.02, SE=0.01, p<0.001), no significant effect of condition (b=-0.02, SE=0.20, p=0.93), and no significant interaction between condition and trial (b=-0.004, SE=0.01, p=0.63). This suggests that the apparent difference in CLASS-based errors across conditions seen in Experiment 1 may have been spurious.

The strong expectation from previous research was that, all things equal, an agglutinating system should be easier to learn than a fusional system. This advantage was not borne out in Experiments 1 and 2. However, we are exploring the early stages of learning these systems, and thus one possibility is that participants in the agglutinating condition were not segmenting the morphemes–i.e., they may have been treating the string of two morphemes as a single morpheme, encoding both NUMBER and CLASS. If so, then we would not expect any difference between conditions. Indeed, the post-test questionnaire reveals that at least some participants failed to segment the stems and morphemes. In Experiment 3 we test whether an advantage for the agglutinating system is revealed if we provide a visual cue to aid segmentation.

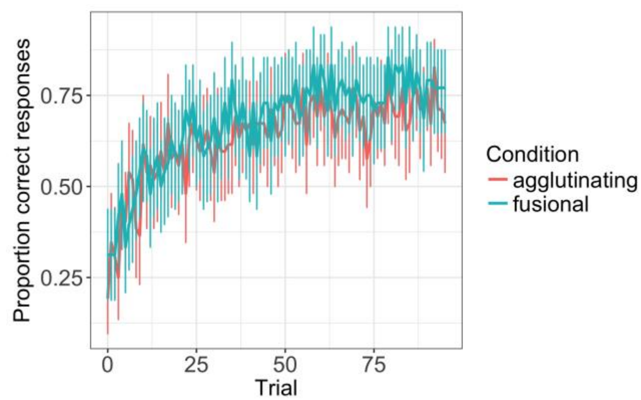

Figure 5: Mean accuracy by trial by condition in Experiment 2.

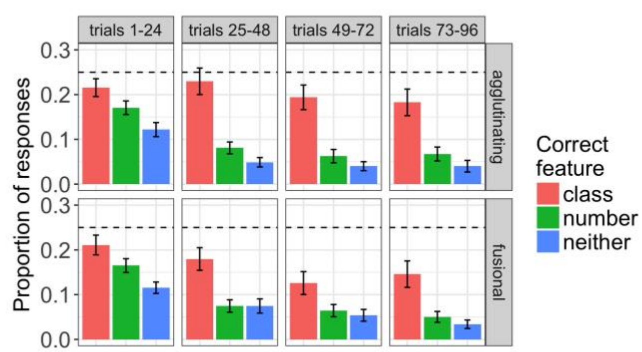

Figure 6: Classification of incorrect choices by correct feature (CLASS correct NUMBER wrong, NUMBER correct CLASS wrong, or neither correct) by trial block by condition in Experiment 2. Trial number is binned for readability. Note that the y-axis range does not display the full range.

Experiment 3

Methods

Participants. 100 participants were recruited on Amazon Mechanical Turk, all self-reported as English native speakers. They were paid $4 for their time. Participants were randomly assigned to one of the conditions (51 in the fusional and 49 in the agglutinating condition).

Materials. The language was identical to Experiments 1 and 2, however, a visual cue to the segmentation of words and morphemes was provided. In each trial, morphemes were highlighted with color (Figure 7). In the agglutinating condition, the CLASS morpheme was highlighted in one color and the NUMBER morpheme in another. Participants were randomly assigned to see either CLASS in orange and NUMBER in blue, or vice versa. In the fusional condition, all four morphemes were randomly assigned a single color so that a participant would either see all morphemes across all 96 trials in orange or all morphemes in blue.

Figure 7: Example trial in Experiment 3, agglutinating condition. This trial shows an inanimate, plural object.

Procedure. The procedure was identical to Experiments 1 and 2.

Results. Figure 8 shows mean accuracy in correct answers across conditions. As in Experiments 1 and 2, participants generally improved from the start to the end as expected, and overall performance appears to be similar across conditions (agglutinating M=0.59, SD=0.25; fusional M=0.59, SD=0.23). We ran a model predicting correct answer by condition, trial (coded 0-95), and their interaction. Condition was dummy-coded, with agglutinating as the reference level. The model revealed a significant effect of trial number (b=0.03, SE=0.005, p<0.001), indicating that participants improved their accuracy over the course of the experiment, but no effect of condition (b=0.23, SE=0.18, p=0.21) and no condition by trial number interaction (b=-0.003, SE=0.01, p=0.64). The latter interaction would have indicated a more rapid improvement in one or the other condition indicating a learnability advantage for one type.

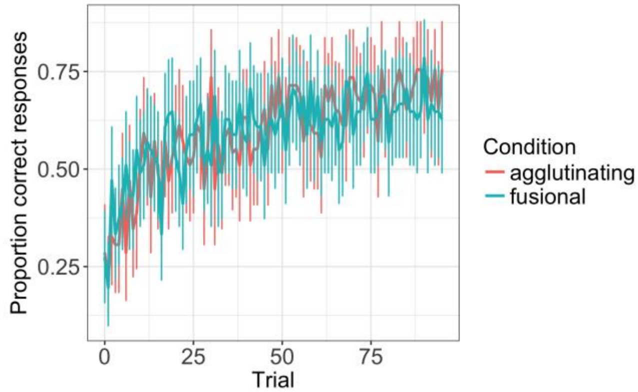

Figure 8: Mean accuracy by trial by condition in Experiment 3.

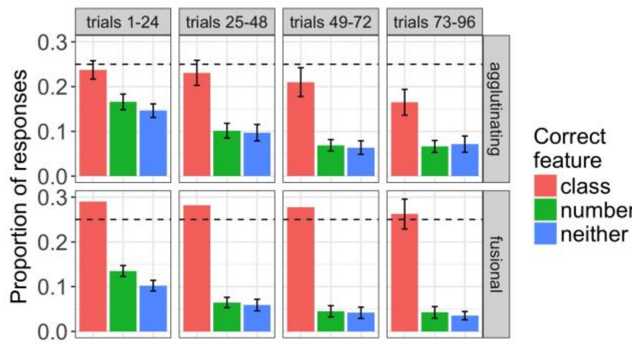

Figure 9: Classification of incorrect choices by correct feature (CLASS correct NUMBER wrong, NUMBER correct CLASS wrong, or neither correct) by trial block by condition in Experiment 3. Trial number is binned for readability. Note that the y-axis range does not display the full range.

We repeated our analysis of error types across conditions (Figure 9). In this case, the model revealed a significant effect of trial (b=0.01, SE=0.004, p=0.03), no significant effect of condition (b=0.29, SE=0.20, p=0.14), and a significant interaction between condition and trial (b=0.02, SE=0.01, p=0.01). Thus, as in Experiment 1, but not 2, participants in the fusional condition were significantly more likely to choose an answer in which only CLASS was correct.

Discussion

It has been claimed that agglutinating systems should be easier to learn because of their inherent transparency: there is a one-to-one mapping between morphemes and meanings in these systems. Here, we directly contrasted a fusional with an agglutinating system, holding the number of morphemes to be learnt constant. We found no clear learnability advantage for agglutinating systems across three experiments. In two of our three experiments, we found a difference in the error patterns between conditions: participants in the fusional condition were more likely to make errors involving NUMBER than CLASS. This error pattern could reflect a bias for transparency. Participants may have been inferring a single four-way distinction, which was possible for CLASS but not NUMBER. However, this effect was not strong enough to result in an overall advantage for the agglutinating system, and it failed to replicate in one experiment. Our results are surprising, given the general and wide-ranging claims in the literature concerning relative ease of learning of agglutinating systems. It may be that an apparent advantage for agglutinating systems reported in the literature is due to confounding differences between the systems in question. However, below we discuss alternative explanations for our failure to uncover the advantage here.

One possibility is that the paradigms we are testing are too small to result in a discernable difference in learnability. It has been noted that compositionality (and therefore transparency) is increasingly beneficial the larger the paradigm is. This was explored computationally by Brighton (2002), who shows that a compositional system for expressing a few features is hardly more learnable than a holistic system covering the same semantic space; the learnability advantage of compositional systems is maximized when each meaning is composed of many binary features. However, the paradigms we used were intentionally small, consisting of just two features, allowing us to test for a learning advantage arising purely from transparency in the absence of benefits associated with increased generalizability. It is therefore possible that the advantages of agglutinating systems derive purely from the fact that they facilitate more rapid generalization, in which case we would not expect to see that advantage in our paradigm.

Another possibility is that agglutinating systems bear additional costs which have not been much discussed in the literature. One such cost is clearly segmentation. Learners can only profit from compositionality if they are able to segment a word into morphemes, but this process might be costly. To eliminate this issue, we used color highlighting in Experiment 3. However, the null effect of condition on overall accuracy was replicated, suggesting that segmentation alone will not suffice to explain why learners did not have an easier time acquiring the agglutinating system.

Compositional structure typically means more material to process for each word: for example, as is typical cross-linguistically, words were longer in our agglutinating condition than in our fusional condition. It is therefore possible that word length (perhaps combined with segmentation cost) has a detrimental effect on the learnability of the agglutinating system.

Finally, seeing a word and its referent (here, an image) in an agglutinating system does not illustrate the meaning of each individual morpheme. Learners of compositional systems need a set of examples to compare and pin down which morpheme expresses which meaning; learning an agglutinating system therefore potentially poses a cross-situational learning problem (similar to that explored by e.g. Yu & Smith, 2007, where multiple words are simultaneously mapped to multiple referents and the precise word-to-referent mapping can only be disambiguated across trials) that is less pronounced for fusional systems. It is possible that this cost, which is often overlooked, together with the length of words and the small size of the paradigm, did not provide a condition under which an agglutinating system becomes easier to learn than a fusional system.

Conclusion

In this paper, we investigated the frequently-made claim that agglutinating systems are easier to learn than fusional systems due to their inherent transparency. Results from three artificial learning experiments did not show the predicted effect. This held even when a visual cue to segmentation was added to help participants discover morpheme boundaries in the agglutinating condition. While some weak evidence for a possible bias for transparent structures was found in participants' error patterns, this did not lead to an overall difference in learning. We argue that this may be due to the small size of the paradigms, which narrow the extent of the benefit for transparency. Some natural language paradigms are of course larger, and these might provide conditions under which the costs of agglutinating systems outweigh those of fusional systems. Setting aside paradigm size, we also argue that agglutinating systems may present additional costs in processing which have not yet been fully explored.

References

Bates, D. M. (2010). lme4: Mixed-effects modeling with R.

Brighton, H. (2002). Compositional syntax from cultural transmission. *Artificial life*, *8*(1), 25-54.

Brown, R. (1973). *A first language: The early stages*. Harvard U. Press.

De Villiers, J. G., & De Villiers, P. A. (1973). A cross-sectional study of the acquisition of grammatical morphemes in child speech. *Journal of Psycholinguistic Research*, *2*(3), 267-278.

Don, J. (2017). What causes languages to be transparent? *Language Sciences*, *60*, 133–143.

Dressler, W. (2003). Morphological typology and first language acquisition: Some mutual challenges. In *Mediterranean Morphology Meetings* (Vol. 4, pp. 7-20).

Dulay, H. C., & Burt, M. K. (1974). NATURAL SEQUENCES IN CHILD SECOND LANGUAGE ACQUISITION 1. *Language Learning*, *24*(1), 37-53.

Goldschneider, J. M., & DeKeyser, R. M. (2001). Explaining the "Natural Order of L2 Morpheme Acquisition" in English: A Meta-analysis of Multiple Determinants. *Language Learning*, *51*(1), 50.

Haspelmath, M., & Michaelis, S. M. (2017). Analytic and synthetic: Typological change in varieties of European languages. In I. Buchstaller & B. Siebenhaar (Eds.), *Studies in Language Variation* (Vol. 19, pp. 3–22). Amsterdam: John Benjamins Publishing Company.

Hengeveld, K., & Leufkens, S. (2018). Transparent and non-

transparent languages. *Folia Linguistica*, *52*(1), 139–175.
Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences*, *105*(31), 10681–10686.
Kirby, S., Tamariz, M., Cornish, H., & Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition*, *141*, 87–102.
Kortmann, B (2005). *English Linguistics: Essentials*. Berlin: Cornelsen.
Slobin, D. (1973). Cognitive prerequisites for the development. *Charles Ferguson and Dan Slobin, Studies in Child Language Development*, 175-208.
Sultana, A., Stokes, S., Klee, T., & Fletcher, P. (2016). Morphosyntactic development of Bangla-speaking preschool children. *First Language*, *36*(6), 637–657.
Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science*, *18*, 414–420.