Extrapolating from reaction times of acceptability judgments to implicit and explicit learning in artificial grammar learning experiments

Tsung-Ying Chen

Department of Foreign Languages and Literature, National Tsing Hua University, Taiwan

Artificial grammar learning (AGL) is an experimental paradigm that briefly exposes learners to language input with hidden regularities and test if learners, without being informed of the learning task, can implicitly (i.e., unconsciously) generalize on target linguistic regularities. The vast majority of AGL experiments recruited adult learners, for whom explicit (i.e., conscious) learning seems to work in parallel with implicit learning. It is thus vital to assess the awareness of acquired linguistic knowledge in adult AGL studies to truly establish the link between learning performance and the acquisition of implicit linguistic knowledge. In this study, we explore the possibility of objectively assessing participants' awareness of acquired linguistic knowledge by analyzing the reaction time (RT) of acceptability judgments in AGL experiments.

With data from three independent AGL studies, we first present evidence that learners respond faster in their acceptability judgments that they have declared as confident than in unconfident judgments. Following Dienes (2007), then, we assume that if a higher confidence level and thus a faster response arises from a conscious application of knowledge, we should observe a *negative* correlation between RT and judgment accuracy (i.e., short RT = high accuracy) if target knowledge is acquired explicitly. Alternatively, no correlation or a positive correlation would suggest the lack of awareness and accordingly, the use of implicit nature of acquired knowledge. In a reanalysis of acceptability judgment data from an additional AGL study, RT demonstrated a *positive* correlation with accuracy for adult learners expected to learn target structural regularities as implicit knowledge. With these findings, we conclude that (i) faster acceptability judgments arise from explicit knowledge is consciously available to guide adult learners to confidently provide rapid acceptability judgments, and (ii) slower acceptability judgments reflect uncertainty, which, if made more correctly, indicate the application of uncontrollable and automatized implicit knowledge. Altogether, we believe that the RT of acceptability judgments in AGL studies could be used as a reliable measure of awareness to probe the implicit/explicit nature of acquired knowledge.

Keywords: artificial grammar learning, acceptability judgment, reaction time, implicit and explicit learning

1. Introduction

In 2004, I was admitted to the graduate program of linguistics at National Chung Cheng University and began my career by working in the Child Language Lab supervised by Professor Jane Tsay. One of the primary ambitious goals in the lab was to construct a child language corpus for Taiwanese Southern Min. Taiwan Child Corpus, or TAICORP (Tsay 2014), to date is still among rare examples offering a sizable amount of valuable child language data of a minor language that could help directly account for child language acquisition and, by extension, advance our understanding of human language. However, since it takes time and tremendous to expand child language corpora, it is worth looking for a different way to test various hypotheses regarding human language acquisition as child language data continue to accumulate.

The search would lead to an experiment approach to the study of language acquisition, namely artificial grammar learning (AGL; Reber 1967).[1] An AGL experiment usually exposes participants to language input that is specifically designed to include a hidden structural regularity (e.g., case marking or word order) in a setting that could minimize conscious noticing of the regular pattern (e.g., by asking participants to memorize input tokens, rather than to explicitly search for rules in them). After the exposure phase, participants will be surprised by an ensuing test for assessing their learning performance, which is usually an acceptability or grammaticality judgment task. As participants are not informed in advance of the purpose of the experiment, the structural regularity in the input, the test of learning performance, and an above-chance test accuracy could be jointly seen as the evidence for implicit (i.e., unconscious) learning of the structural regularity.

With the flexibility in its experimental design, the AGL paradigm could be implemented to study issues that are not or cannot be addressed in child language research. For instance, one could run AGL experiments to test the learnability of particular structural regularities in an attested language while we are waiting for linguistics to build the child corpus of the language. In addition, as the training input in AGL studies could be freely manipulated to comprise all logically possible structural patterns, AGL studies could be adopted to test if untested languages could be acquired by human learners. AGL studies may, to some extent, be more convenient than child language

[1] Note that in this paper, *acquisition* and *learning* are used interchangeably and do not respectively imply L1 or L2 attainment as they usually do in the literature. In particular, AGL has been adopted to study language acquisition in early infancy, and L1 and L2 could be attained presumably via both implicit and explicit learning (see also fn.2).

studies for linguists to distinguish learnable from unlearnable languages and explain the limits of our language faculty and typological asymmetries. That is, distributional asymmetries or gaps in language types might be attributed to intrinsic learning biases in favor of specific structural patterns in our innate linguistic knowledge (see §3 for examples). Implications for the awareness of target knowledge in the findings of AGL studies further enables the possibility to answer research questions about the process of L1 acquisition and the nature of L1 knowledge, which are believed to be largely unconscious in early infancy and childhood (Ellis 2008).[2]

One potential issue in AGL studies that could nevertheless undermine conclusions regarding the implicit learning of hidden structural regularities is that most AGL studies have recruited adult participants as learners, for whom explicit (i.e., conscious) learning seem to work in parallel with implicit learning (e.g., Hulstjin 2005). It is thus necessary to disentangle the effects of implicit and explicit learning on adult participants' performance in AGL studies, so the performance arising from the application of explicitly acquired knowledge is not mistaken as evidence of implicit and perhaps innate linguistic knowledge. The issue is particular critical for AGL studies that seek to explain an asymmetry in the learnability of structural patterns with innate and unconscious learning biases (see Culbertson 2012 and Moreton & Pater 2012a, b for reviews). To the best of our knowledge, Moreton & Persova (2016), Pertsova & Becker (2021), and Chen (2022a) are the only examples that have emphasized the distinction between implicitly and explicitly acquired knowledge in the investigation of intrinsic inductive biases using the AGL paradigm.

To investigate the implicit/explicit nature of acquired knowledge, an AGL experiment must measure learners' awareness of the knowledge. Our goal in this study is to explore the reaction time (RT) of acceptability judgments in AGL studies as a reliable measure of awareness. We will reanalyze experimental data from independent AGL studies by correlating RT with confidence, which at least partially reflects the awareness level, and then with judgment accuracy. We will aim to validate RT as a measure of awareness by testing if the RT-confidence-accuracy correlation varies significantly by the application of implicit and explicit knowledge.

In §2, we will begin with a discussion of the connection between confidence and learners' awareness of linguistic knowledge and how RT could more *objectively* reflect the link and serve as a more reliable awareness

[2] As Hulstijn (2015) rightly suggested, however, it should not be taken for granted that all the structural aspects in L1 are acquired before learners are made explicitly aware of the language learning process.

measure. In §3, we will review individual AGL studies that are the sources of our data used to establish the correlation between RT, confidence, and judgment accuracy. The validation process and the data analysis will be both illustrated in §4, and the findings will be discussed in §5.

2. Reaction time as a confidence-based awareness measure

Among previously proposed awareness measures, Dienes' (2007) confidence-based *zero-correlation criterion* has been commonly adopted in AGL studies with an acceptability judgment task (e.g., Chan & Leung 2014, 2018; Graham & Williams 2018; Chen 2022a). In these studies, learners in a test session are asked to rate their confidence level after providing each acceptability judgment. If there is a positive correlation between learners' judgment accuracy and their *subjective* confidence level (i.e., confident = accurate), the learners could be in a conscious state when they extend acquired linguistic generalizations to their acceptability judgments. If there is no correlation between an above-chance judgment accuracy and confidence ratings, we would conclude that learners are not aware of acquired target knowledge. Finally, if there is a *negative* correlation, it could be that implicitly acquired knowledge is superior to explicitly acquired knowledge (Dienes 2007:57).

Maie & DeKeyser (2020) nevertheless criticized the use of subjective confidence ratings as an awareness measure, claiming that adult learners might be too humble to rate themselves as confident even if they are aware of acquired knowledge to some extent. Instead, Maie & DeKeyser proposed to measure learners' awareness in an online task in which real-time application of implicit/explicit knowledge could be directly observed in the variation of RT. In their AGL study focusing on the learning of syntactic knowledge, Maie & DeKeyser followed Granena (2013) and tested learners with an audiovisual word-monitoring task. This task required the learners to respond as quickly as possible to a target word in each test sentence, which would immediately follow a grammatically legal or illegal string (e.g., Tom thinks Mary *is* angry vs. Tom thinks Mary **are* angry). Successful detection of grammatically illegal strings in online processing would delay participants' responses to upcoming target words. Slower RT could thus be viewed as evidence showing the attainment and application of implicit and highly automatized linguistic knowledge. Conversely, insensitivity to ungrammatical strings would not delay learners' responses to upcoming target words, possibly because learners do not have the essential implicit knowledge that can be spontaneously deployed in online processing. As RT is unlikely to be swayed by subjective factors that are not directly related to the level of knowledge awareness, RT

measured in the word-monitoring task is a potentially informative *objective* measure of adult learners' implicit structural knowledge. Maie & DeKeyser compared analyses based on RTs and subjective confident ratings and found that the two measures of awareness pointed to different types of acquired knowledge; only the subjective measure suggested the implicit learning of target knowledge. The mismatch has led Maie & DeKeyser to emphasize the risk of misinterpreting learner's performance with a subjective measure of awareness and the importance of adopting an objective measure of awareness.

The proper application of the word-monitoring task in other domains (e.g., phonology) awaits further investigation, so we turn to explore the possibility of measuring adult learners' awareness with RT from acceptability judgment tasks, which have been the most common assessment of learners' performance in adult AGL studies. Attempts to assess learners' awareness level via the response latency of acceptability judgments in AGL studies are rare, if not completely absent, perhaps owing to the ambiguity of interpreting the source of slow or fast judgments. For example, learners could either respond more rapidly when they are explicitly aware of the patterns or more slowly when it is more effortful to deploy explicit knowledge (Ishikawa 2019: 1384-1385). This ambiguity could be solved by triangulating the relationship between RT, subjective confidence level, and the nature of acquired linguistic knowledge in AGL studies, which is the goal of the current study.

With data from independently ongoing or completed AGL studies, we first attempt to correlate the RT of acceptability judgments to adult learners' subjective confidence level. Once we are able to map RT to the degree of confidence (see §4.1), we could use RT as a fine-grained confidence-based awareness measure that is not dependent on the subjective reflection on one's own confidence. Then, following the zero-correlation criterion, we would be able to validate RT as an awareness measure by correlating RT to judgment accuracy. For instance, assuming that faster judgments reflect, in part, a higher level of confidence, which in turn largely depends on conscious awareness of linguistic knowledge, the zero-correlation criterion predicts a *negative* RT-accuracy correlation (i.e., shorter RT = higher accuracy) for linguistic patterns that adult learners acquire as *explicit* grammatical knowledge. By contrast, no negative RT-accuracy correlation would be found for grammatical generalizations acquired and applied implicitly by adult learners. Alternatively, if more confident responses are slower, a *positive* RT-accuracy correlation (i.e., longer RT = higher accuracy) would be found only for adult learners acquiring *explicit* knowledge according to the zero-correlation criterion.

3. Data sources for analyzing correlations between reaction time, confidence, and judgment accuracy

In the current research, we will first combine data from three AGL studies and analyze the general RT-confidence correlation (Chen 2022a, b, c). The three studies have a focus on the learning of different grammatical aspects but share a similar experimental design, which will be reviewed in §3.1–§3.3. In the training phase, adult learners were exposed to aural or visual input and instructed to try their best memorizing the input. After the training phase, one or more test phases followed, in which learners were asked to give binary acceptability judgments (Yes vs. No) of whether test items conform to the hidden linguistic regularities in the training input, and RT in milliseconds was measured for each judgment. After each judgment, learners were asked to reflect on their judgment and rate their binary confidence level (Yes vs. No) so the data allow us to correlate RT with confidence. After confirming a close RT-confidence correlation, we will extend our findings to an additional AGL study (Chen 2020; see §3.4) to test if the RT-accuracy correlation varies by the implicit/explicit learning of grammatical knowledge. The core experimental design of Chen (2020) is similar to that of Chen (2022a, b, c), except that learners did not provide any subjective confidence level for their acceptability judgments in the test phase. Without any other measure of awareness, the data from Chen (2020) are suitable for demonstrating how an RT-accuracy correlation could help validate the claim of implicitly and explicitly learning in AGL studies.

3.1 An inductive bias against adjacent tonal levels

The first AGL study we will include to investigate the RT-confidence correlation is Chen (2022a), which addressed the long-standing debate over whether contour tones are merely phonologically represented as sequences of tonal levels or could also be represented as single constituents.

Whether contour tones can form a unit or not has practical implications for the projected typology of tone languages. In particular, the phonological representation of tones determines how the Obligatory Contour Principle (OCP; Leben 1973) is applied to ban adjacent tonal sequences. If contour tones are merely presented as a sequence of tonal levels, we would expect only languages in which adjacent identical tonal levels are avoided, whether they are part of a contour tone or not (e.g., *H<u>L</u>-<u>L</u>H, *<u>H</u>-<u>H</u>L, *H<u>L</u>-<u>L</u>). This type of tone languages is common. Alternatively, if contour tones also form single units, the OCP could be extended to ban adjacent identical tones

as well, whether they are level or contour tones (e.g., *H-H, *HL-HL, *LH-LH). There are nevertheless three primary issues with the unit-based view on the representation of tones. First, tone languages that are claimed to support the unit-based view are usually Chinese languages, whose modern tonal system has undergone complex diachronic tonal changes. Therefore, tonal patterns found in the system does not necessarily represent synchronic phonological processes. Second, the unit-based view on tonal patterns could be reanalyzed with a non-unit-based approach (e.g., Chen 2010). Finally, the unit-based view overgeneralizes on the type of tone languages, as a language that bans same adjacent terminal tonal levels as well as adjacent tones has yet to be discovered (Wee 2019: 167–168).

Chen (2022a) attempted to solve the debate using the AGL paradigm with the hypothesis that if tones could be represented as single constituents, the OCP banning adjacent identical tones (OCP-Unit) would be as learnable as the OCP banning only adjacent terminal tonal levels (OCP-Terminal). In Exp I, 90 adult participants speaking Taiwan Mandarin as their L1 were recruited and randomly assigned to the OCP-Unit, OCP-Terminal, and Control groups. The first two groups were exposed to disyllabic tonal patterns created with H, LH, L, and HL without violations of the two respective constraints. The Control group was exposed to all possible di-tonal combinations of the four tones except L-L (i.e., Third-tone sandhi violation) and was not expected to learn any tonal pattern from the input. In an acceptability judgment task that was administered immediately following the exposure phase, all three groups were asked to provide binary judgments of whether nonwords consisting of a new set of segments and all possible di-tonal combinations except L-L were acceptable. After each valid response, the participants were also asked to rate their confidence on their judgment.

The analysis of the results indicated that while both of the OCP-Unit and OCP-Terminal groups demonstrated the acquisition of the target tonal patterns, their confidence ratings were positively correlated with their judgment accuracy. However, if only the subset of unconfident responses was analyzed, there was only a sign of learning OCP-Terminal but not OCP-Unit. The same analysis also showed that the OCP-Unit group did not outperform the Control group in terms of detecting the violation of OCP-Unit in test items. Finally, whereas no participant in the OCP-Terminal group could explicitly verbalize the tonal patterns hidden in the training input, a few participants of the OCP-Unit group were able to describe the avoidance of same tones in the training input. Taking all the evidence into consideration, Chen (2022a) concluded that OCP-Terminal but not OCP-Unit is part of the implicit linguistic knowledge, and tones might not be encoded as single constituents at the level of abstract phonological computation.

3.2 An inductive bias favoring opaque vowel harmony

In another AGL study that measured learners' awareness with subjective confidence ratings in acceptability judgment tasks, Chen (2022b) revisited a potential difference in the learnability of opaque and transparent vowel backness harmony (VH; e.g., Gafos & Dye 2011) investigated in Finley (2015).

In VH, a vowel agrees with the previous vowel in backness, and the agreement applies directionally through the VH domain (e.g., from left to right: /$V_{+bk}V_{-bk}V_{-bk}$/ → [$V_{+bk}V_{\mathbf{+bk}}V_{\mathbf{+bk}}$]). However, not all the vowels play the same role in VH; some vowels block the backness agreement (e.g., /y/ and /ɛ/ in Hungarian), while others neither block nor participate in the agreement (e.g., /i/ in Hungarian). These vowels are known as *neutral* vowels, which are either *opaque* or *transparent* to a VH process (opaque: /$V_{+bk}y_{-bk}V_{-bk}$/ → [$V_{+bk}y_{-bk}V_{-bk}$]; transparent: /$V_{+bk}i_{-bk}V_{-bk}$/ → [$V_{+bk}i_{-bk}V_{\mathbf{+bk}}$]).

Finley (2015) argued that learners would acquire opaque VH more easily than transparent VH for two reasons. First, there would be an intrinsically smaller subset of input that could serve as the positive evidence for transparent VH. For instance, when a transparent neutral vowel, such as /i/, is enclosed in a pair of front vowels in the output (e.g., [e-i-e]), there is no indication whether /i/ participates in VH or is transparent to VH. Second, a constraint-based analysis suggested that more constraint rankings favor outputs with opaque VH over those with transparent VH. In other words, by pure chance, it would be easier for learners to reach the constraint ranking that only allows for the production of opaque VH outputs. In a series of AGL experiments, Finley discovered that opaque VH was indeed learned more efficiently by adult subjects, and only one experiment demonstrated the successful learning of transparent VH when input tokens were doubled. Both findings are in line with Finley's hypotheses.

Finley's account of the learning bias nevertheless faces a challenge from a theoretical perspective. That is, the number of constraint rankings predicting opaque and transparent VH outputs could vary depending on the constraints used to account for the two VH processes. In an account that predicts no difference in the probability of converging on the two VH grammars, one could claim that the observed learners' performance is influenced by extragrammatical factors. Or, there could exist an account in which chance is higher for learners to converge on a transparent VH grammar.

Given the above concern, Chen explored the potential inductive bias against transparent VH from a theory-neutral cognitive perspective with methodological improvements. Specifically, Chen tested if the *starting-small* effect (e.g., Newport 1990; Lai & Poleitek 2013) is the foundation of the

inductive bias. The starting-small effect is grounded on the limited cognitive capacity that is presumably optimized for processing and memorizing the structure of more frequent but smaller constituents in a naturally skewed distribution (i.e., Zipf's Law). The base knowledge formed in the processing of the smaller constituents would in turn support structural parsing of larger but less frequent constituents. The effect may strengthen the learning of local VH that can be directly observed from phonologically shorter words (e.g., disyllabic stems), and the attained base knowledge could further facilitate the learning of local and opaque VH on phonologically longer words (e.g., suffixed forms). Since the learning of transparent VH is only observed on longer words, it is thus not supported by the base knowledge. In sum, it is assumed that the starting-small effect underpins the development of the implicit grammatical knowledge of opaque VH and the difficulty of learning transparent VH.

To test this hypothesis, Chen recruited a total of 98 adult participants speaking Taiwan Mandarin as their L1 in two experiments and randomly assigned them to two training conditions. In the Balanced and Starting-small condition, participants were all exposed to randomly presented individual (rather than paired) CV_1CV_2C stems and their CV_1CV_2C-V_3 suffixed forms. The stems were manipulated to include an opaque V_2 (opaque stem), a transparent V_2 (transparent stem), or a non-neutral front/back V_2 that agrees with V_1 (harmonized stem). The backness of V_3 in suffixed forms was determined based on V_1 when V_2 is transparent to VH, or based on V_2 in the other two cases. The differences between the two learning conditions lay in how input tokens were presented. In the Balanced condition, disyllabic stems and their suffixed forms were equally frequent, and their presentation was entirely random. In the Starting-small condition, disyllabic stems were three times as frequent as their suffixed forms and were always presented before the suffixed forms. Following the training session was an immediate and a delayed test session that presented pairs of suffixed forms in which the V_3 may agree with V_1, V_2, or neither of them. Participants were prompted to choose the more acceptable suffixed form and then rate their confidence in their choice (Yes vs. No).

The results focusing on judgments of suffixed forms with a *novel* stem elicited from the immediately test session are illustrated in Figure 1, which compares the judgment accuracy rates across the two groups (Group) with regard to their choices of V_3 based on different stem types (Stem Type) and the confidence rating of the judgment. For both groups, the learning performance was seemingly better when the learners rated their judgments as confident, which suggests the awareness of target VH patterns to some extent. However, the implicit knowledge that arose in participants also seems to differ

across the two groups based on the patterns found with the unconfident judgments. The Starting-small group still demonstrated a high judgment accuracy for opaque VH when the learners were unconfident, which is an indicator of learning opaque VH as implicit knowledge according to the zero-correlation criterion (i.e., acceptability judgments are accurate whether learners are confident or not). By contrast, the Balanced group demonstrated a chance-level accuracy for all three VH patterns in their unconfident judgments; there was no sign of acquiring any implicit VH knowledge for the Balanced group. These results allow Chen to argue in favor of the cognitive account of the inductive bias against transparent VH.

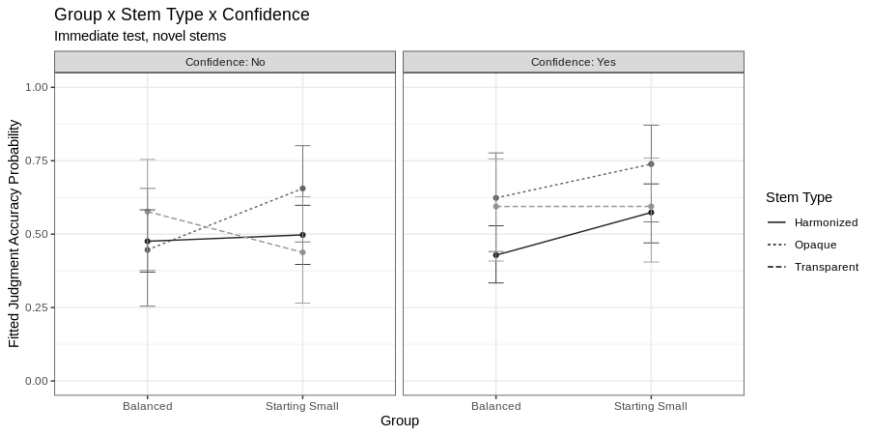

Figure 1. The comparison of VH judgment accuracy for novel suffixed forms across Group, Stem Type, and Confidence in Chen (2022b)

3.3 The learning of hidden orders in letter strings

The last AGL study (Chen 2022c) that supplies the data for our cross-experimental analysis of the RT-confidence correlation is a partial replication of Reber's (1967) seminal AGL study testing the implicit learning of hidden patterns in letter strings. In Reber's original study, adult participants were trained with letter combinations of T, P, X, V, and S, with orders either generated systematically by a finite-state machine (Target) or generated randomly (Control). After completing the training session, the learning performance was assessed in a recall task and an acceptability judgment task with novel TPXVS strings, and the Target group outperformed the Control group in both tasks.

In the ongoing study, Reber's experimental design was partially replicated in Chen (2022c) with the identical finite-state algorithm for generating regular letter strings. However, to test if learners can generalize on abstract patterns rather than memorize and reuse letter chunks, Chen exposed participants to TPXVS and ROBYL strings. In the test sessions, strings were also generated with a different set of letters, namely UICZN. This way, an above-chance and significantly better test performance of the Target group would be viewed as stronger evidence for implicit rule abstraction.

The study has so far recruited 42 adult participants to complete the experimental online via their desktop or mobile device in ENIGMA (https://lngproc.hss.nthu.edu.tw/ENIGMA), a Web-based platform for running large-scale online AGL experiments. During the training session, structured or randomly ordered strings were presented visually in random order at the center of their device. When the training session was over, participants immediately judged if a new set of strings followed the patterns hidden in the training input. A delayed session was administered to evaluate the lasting effect of acquired target knowledge. As with the other two AGL studies reanalyzed in this paper, we only discuss data collected in the immediate test session for the consistency in our comparison.

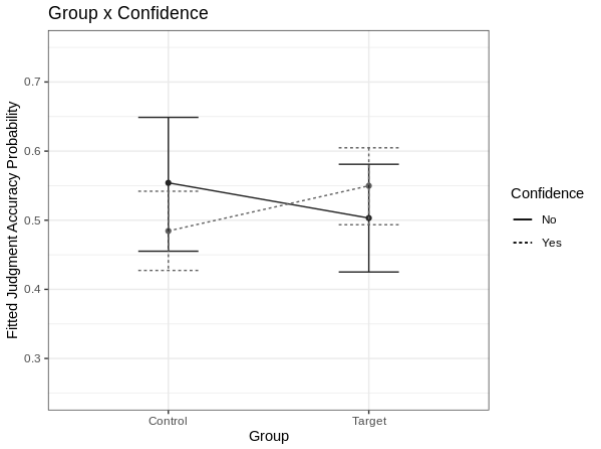

Figure 2. The between-group comparison of judgment accuracy by confidence level in Chen (2022b)

The judgment patterns of the subset are visualized in Figure 2 with a between-group comparison of judgment accuracy by subjective confidence level. The figure suggests a potential interaction between Group and Confidence as a positive correlation between confidence level and judgment

accuracy for the Target group. In addition, above-chance performance can only be found with confident judgments for the Target group. These findings contradict the claim in Reber (1967) that structural regularities in ordered strings generated with a finite-state machine are generalized implicitly, which further highlight the importance of awareness measures in AGL experiments.

3.4 An inductive bias against non-final rising tone

The AGL study that offers an opportunity to test the difference in the RT-accuracy correlation across learning conditions is Chen (2020), in which learners were assumed to have an intrinsically implicit learning bias in favor of phonetically natural tonal phonotactics.

In previous cross-linguistic research of the distribution of tones, it has been found that phonetically longer tones are more likely to be banned in non-domain-final positions (Zhang 2002; 2004; 2007). Crucially, many languages prohibit non-final rising tones, presumably because (i) a rising contour needs more time than a falling contour or a flat pitch level to be fully realized and (ii) non-final syllables are intrinsically shorter than final syllables. The avoidance of the phonetic mismatch between tones and their hosts also significantly influences the productivity of partially lexicalized tone sandhi patterns, as indicated in a series of experimental studies focusing on Standard Mandarin, Taiwanese Southern Min, and Tianjin Chinese (i.e., Zhang & Lai 2010; Zhang et al. 2011; Zhang & Liu 2016). The consistent finding in these studies is that tone sandhi patterns are more productive and extend to nonce words more easily if they are phonetically natural (and/or have a higher type frequency). With this evidence demonstrating the strong effect of phonetic naturalness in tonal phonology, the question is raised whether phonetic naturalness is part of the intrinsic implicit knowledge of tones and plays a critical role in the learning of tonal patterns. One would assume that, if phonetic naturalness guides the learning of tonal phonology, learners of a tone language should be biased to acquire tonal phonotactics that have a clear articulatory or perceptual foundation.

This hypothesis was tested in Chen (2020) in an AGL paradigm, in which 48 adult participants speaking Taiwan Mandarin as their L1 were exposed to disyllabic input created with tones H, LH, L, and HL without either non-final rising (*NonFinalR) or non-final high-level tones (*NonFinalH) in Exp I. The constraint banning non-final high-level tones was assumed to be less learnable as it is neither structurally simpler nor phonetically natural.[3]

[3] See Pater & Moreton (2012a, b) for a review of phonological inductive biases toward structurally simple and phonetically natural patterns.

Chen also excluded phonemic retroflex consonants [ɹ, ʂ, tʂ, tʂʰ] in Taiwan Mandarin from the onset position in all input tokens to create a segmental gap that was expected to be a learnable contextual generalization for both groups. The absence of retroflexes served as the baseline for comparison since if the tonal constraints are part of our implicit grammatical knowledge, they should be as learnable as the segmental generalization.

In this experiment, the learning performance was tested in an immediate and delayed acceptability judgment task, in which the participants were asked to judge if test items created with a new set of disyllabic segmental and tonal combinations were acceptable (i.e., Yes vs. No), and RT was measured in milliseconds. The analyses of the results suggested that the *NonFinalR group outperformed the *NonFinalH group with an above-chance judgment accuracy in terms of whether the test items violated the respective tonal constraints. In addition, the tonal and segmental constraints were learned equally well only for the *NonFinalR group. Along with evidence from another AGL experiment, Chen concluded that there is an inductive bias toward the learning of phonetically natural tonal phonotactics.

The findings in the experiment are nevertheless open to an alternative, non-grammar-based interpretation for two reasons. First, the presence of retroflex onsets in the test items could be detected by *consciously* comparing the test items to the training tokens retrieved from the phonetic memory. Since the learning of no retroflex onset is not necessarily grammatical, the similar performance in the learning of *NonFinalR neither implies implicit grammar learning of the target tonal phonotactics. In addition, the learning of *NonFinalR could itself be attributed to a higher perceptual sensitivity to tonal contours for native speakers of a language with a rich contour tone system (e.g., Gandour 1981; 1983). In other words, the subjects might have found it easier to *explicitly* detect a non-final rising tone than a non-final high-level tone in the test items, as the former is perceptually more salient. Thus, to answer the question whether the adult subjects in Chen (2020) did indeed acquire *NonFinalR and the segmental phonotactics as implicit grammatical generalizations, we need a proper awareness measure. With a link established between RT and learners' confidence, we will be able to revisit the findings in Chen (2020) and attempt to elucidate the implicit/explicit nature of the acquired knowledge.

4. Data analysis

As planned, our data analysis will begin by establishing the connection between RTs and confidence ratings in §4.1, which would first map

faster/slower judgment RT to binary subjective judgment confidence levels. This link would then represent a continuous *objective* scale reflecting the level of confidence, which would be used to infer the implicit and explicit nature of acquired knowledge in the analysis of the RT-accuracy correlation in §4.2.

4.1 Reaction time and subjective confidence rating

For the analysis of the RT-confidence correlation, the grand data set with a total of 8,190 acceptability judgments was merged from the three data subsets from the immediate test session in Chen (2022a, b, c) reviewed in §3.1–3.3.[4] RTs in milliseconds were log-transformed (cf. Brysbaert & Stevens 2018) and *z*-scored within subjects to accommodate inter-subject and cross-experimental variation. Judgments with an outlier RT ($z > 2$ or < -2) were removed from the grand data set (N = 216; 2.6%).[5] The remained 7,974 judgments were submitted to a linear mixed-effects model in R 4.1.2 (R Team Core 2021) with the *lme4* package (Bates et al. 2021), in which RT was regressed against the dummy-coded fixed predictor Confidence (Yes vs. No) with the by-subject random intercept taken into consideration.

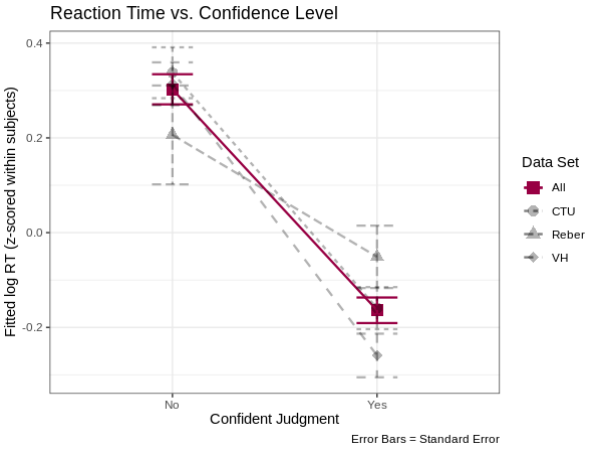

Figure 3. The relationship between RT and subjective confidence level; All = the grand data set; CTU = Chen (2022a); VH = Chen (2022b); Reber = Chen (2022c)

[4] Chen (2022a) = 4,241; Chen (2022b) = 2,887; Chen (2022c) = 1,062.
[5] Chen (2022a) = 113 (2.7%); Chen (2022b) = 66 (2.3%); Chen (2022c) = 27 (2.5%).

Confidence was found to have a significant effect on RT as confident judgments were noticeably faster than unconfident ones (β = –0.466, *SE* = 0.021, $t(1145.8) = -22.42$, $p < .001$). The significant effect is consistent across the three data subsets submitted to the same linear mixed-effects model as illustrated in Figure 3.[6] Thus, it is safe to conclude that faster judgments are also more confident judgments, at least in these three AGL studies.

4.2 Reaction time and judgment accuracy

Following the above analysis, we revisit the findings in Chen (2020) and aim to reach a conclusion with regard to the implicit and explicit nature of acquired tonal and segmental generalizations. Crucially, if the detection of rising/high-level tones and retroflex segments in the test session arises from an explicit comparison between test items and the phonetic memory of training items, we would find a *negative* correlation between RT and judgment accuracy (i.e., shorter RT (more confident) = more accurate). Alternatively, the implicit nature of the target knowledge would be indicated by a *positive* RT-accuracy correlation (i.e., longer RT (less confident) = more accurate) or no correlation at all. As reviewed in §3.4, Chen assumed that *NonFinalR and the segmental phonotactics (i.e., no retroflex onset) are phonological generalizations learnable via the intrinsic and implicit linguistic knowledge, whereas *NonFinalH is not. Accordingly, we should not discover a non-negative RT-accuracy correlation in the successful learning of *NonFinalR and the segmental phonotactics.

In our analysis, we first took a subset of 3,072 trials from the immediate test session in Chen's Exp I,[7] which composed only of test items including a non-final rising tone, a non-final high-level tone, or a retroflex onset. The 49 adult learners should thus have rejected these test items as correct judgments. With this subset, we log-transformed RT in milliseconds and *z*-scored log-RTs within subjects as in §4.1. We also excluded 120 (3.9%) judgments with an outlier RT ($z > 2$ or < -2). The remaining 2,952 acceptability judgments were submitted to a logistic mixed-effects model in R 4.1.2 with the *lme4* package, in which the probability of correct rejections was regressed against RT, Group (*NonFinalH vs. *NonFinalR), and the type of violated phonotactics (Segment vs. Tone) with a three-way RT × Group ×

[6] Separate analyses suggested no significant between-group difference in the negative Confidence effect on RT except in Chen (2022a), where the negative effect was found to be smaller for one of the three learner groups. Accordingly, the main effect of Confidence is generally consistent despite different learning input and conditions in each study.
[7] The data set was retrieved from https://osf.io/k36qx/ on Mar 1, 2022.

Type interaction. By-subject and by-item intercepts as well as the by-subject slope or RT were included following the model selection process in Matuschek et al. (2017).

The model is visualized in Figure 4 with the three-way interaction between the independent variables. Crucially, there was an opposite effect of RT on the probability of correct rejections across the two groups of adult subjects. For those assigned to the *NonFinalH group, there was either an inverse or no RT-accuracy correlation. When the test items were presented with a retroflex onset, the probability of rejection was higher with faster and thus, by hypothesis, more confident judgments. When the test items came with a non-final high-level tone, there was no difference in the probability of rejection for fast and slow judgments. It is also important to note that the rejection of test items violating the target tonal phonotactic was only around the chance level. For the subjects of the *NonFinalR group, the effect of RT was always positive, as slower and possibly less confident judgments led to more correct rejections, regardless of whether test items had a non-final rising tone or a retroflex onset. This between-group difference in the RT-accuracy correlation was reflected in a marginal two-way RT × Group interaction in the mixed-effects model (β = –0.132, SE = 0.069, z = –1.919, p = .055).

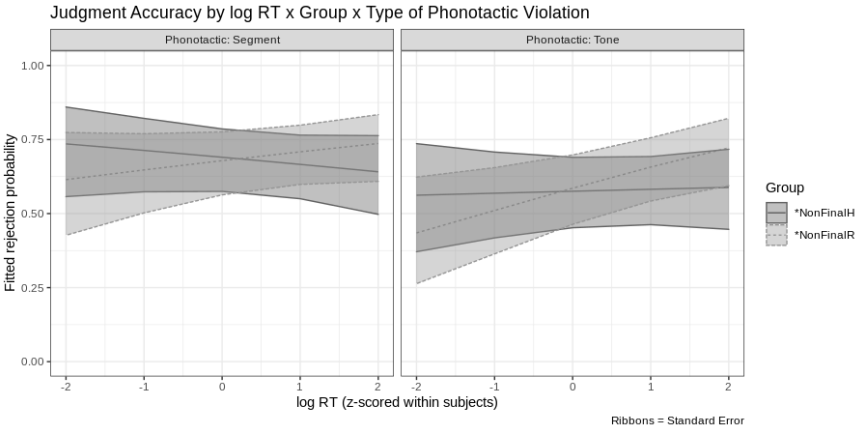

Figure 4. The accuracy of rejecting test items violating the target segmental/tonal phonotactics by log-RT across the two learner groups in Chen (2020).

As we have found a distinctive RT-accuracy correlation for the two learner groups, we can further expand our analysis to cover the judgments from the delayed test session to test the lasting effect of learning target tonal

phonotactics, which was observed only for the *NonFinalR group in the original study. If this effect arose from implicit grammar-based learning, we would expect to discover a similar positive RT-accuracy correlation only for the *NonFinalR group in the delayed test session, too.

The delayed test session in the original study included 3,072 judgments, which were paired with those from the immediate test session and were meant to be rejected with a successful learning of the target segmental and tonal generalizations, too. Following the same RT transformation and screening procedure, we dropped 130 (4.2%) judgments with an outlier RT. The other 2,942 judgments were submitted to the same logistic mixed-effects model adopted in the previous analysis, and the three-way RT × Group × Type interaction is visualized in Figure 5. The model suggested a consistent positive RT-accuracy correlation for the *NonFinalR group with no regard to the violation of target phonotactics, which would be the primary source of the significant main RT effect (β = 0.166, SE = 0.065, z = 2.539, p = .011), although the between-group difference in the RT effect was also marginal (β = –0.115, SE = 0.065, z = –1.760, p = .078).

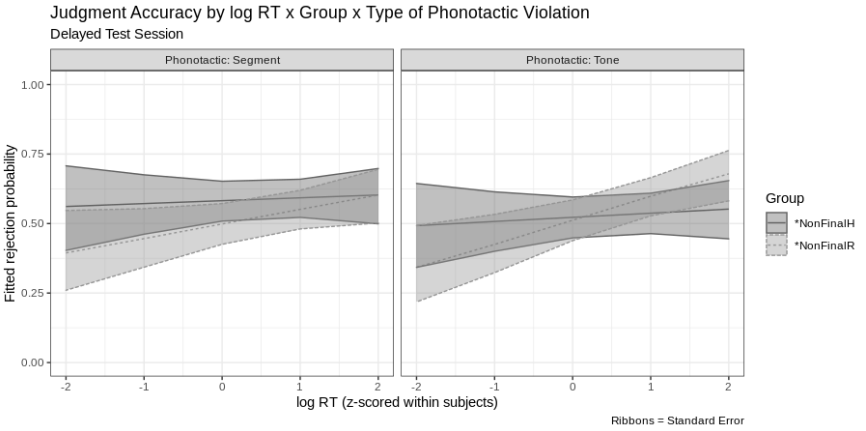

Figure 5. The accuracy of rejecting test items violating the target segmental / tonal phonotactics by log RT across the two learner groups in the delayed test session in Chen (2020).

5. General discussion

In §4.1, we provided evidence showing an inverse correlation between RT and subjective confidence ratings with acceptability judgment data from three theoretically unrelated but methodologically similar AGL studies. This

evidence set the stage for testing the predictions that there is a correlation between RT and judgment accuracy and that this correlation is dependent on whether learning is expected to be implicit. In §4.2, our analyses of data from a separate phonological AGL study (i.e., Chen 2020) helped strengthen the original conclusion that there exists an implicit grammar-based inductive bias for learning a phonetically natural generalization. Notably, we found that the *NonFinalR learners exposed to the phonetically natural tonal patterns rejected test items more correctly when they hesitated and thus possibly lacked confidence. The finding was consistent in an additional analysis of data from a delayed test session, which indicated a lasting effect of implicit learning in AGL settings (see also Martin & Peperkamp (2020)). By contrast, the *NonFinalH group exposed to a phonetically unnatural tonal pattern not only failed to extend the pattern (i.e., more unlikely to correctly reject novel test items violating it), but this group was also more inclined to correctly detect a retroflex onset when judgments were faster rather than slower. In other words, the learners might have been explicitly aware of the absence of retroflex consonants in the training input, even if the generalization could have been implicitly acquired on a grammatical basis.

The above findings should be compelling evidence for the use of RTs measured from acceptability judgments in AGL experiments as a confidence-based index of knowledge awareness. There are nevertheless still some residual issues related to the continuous awareness measure to be discussed before closing. First, our inference of the RT-accuracy correlations found in §4.2 is rooted in Dienes' proposal of the zero-correlation criterion, which treats the lack of positive confidence-accuracy correlation as the evidence for implicit learning. However, logically speaking, the lack of a positive confidence-accuracy correlation could be at best interpreted as no evidence for explicit learning, which may be indeed due to the dominance of implicit learning or due to a failure to detect the effect of explicit learning. Thus, a stronger version of the original zero-correlation criterion is an *inverse-correlation criterion* that only a negative confidence-accuracy correlation could be truly viewed as solid evidence of implicit learning, as it could imply the dominance of implicit knowledge (Dienes 2007:57). When we adopt RT as the awareness measure, implicit learning means a higher judgment accuracy for slower responses, which is exactly what we have discovered earlier in §4.2. In future studies, it might thus be worth examining if RT, as a more fine-grained awareness measure, is more sensitive than subjective confidence ratings to the inverse correlation.

The indirect link between fast responses and the application of explicit knowledge via subjective confidence ratings might seem tenuous to some readers as well, since explicit knowledge is usually assumed to be less

automatized than implicit knowledge, and a controlled process is deemed to slow rather than fast. However, since conscious knowledge could be automatized as well (e.g., Maie & DeKeyser 2020), rapid responses are also possible outcomes of explicit knowledge application. This knowledge might arise from explicit search for possible (sub)regularities from learning input by adult learners and is more dominant as an initial heuristic in making acceptability judgments, especially for those who have rich experience with explicit learning (e.g., Lichtman 2013). When test items do not contain grammatical aspects that clearly contradict or conform to the conscious knowledge, adult learners cannot be guided to make quick decisions. This is when implicitly developed knowledge, if any, to gradually take over the decision-making process and unconsciously drive adult learners to give accurate but slower judgment. The complete model of implicit and explicit knowledge learning and application is very likely to be more complex and dynamic than the one depicted above. On one hand, there might be no clear cut-off between the use of explicit and implicit knowledge (e.g., Dienes & Perner 1999). On the other, there could be different forms of implicit and explicit knowledge (e.g., more vs. less automatized explicit knowledge) that would possibly compete with each other to become more dominant along the time scale until learners settle on a final judgment. Assuming this is true, the continuous RT measure obtained from acceptability judgments might be more powerful for modeling the chronological process of knowledge retrieval and application in AGL settings, a hypothesis that warrants a careful inspection with more data.

6. Conclusion

In the current study, we have mapped RT measures of acceptability judgments from three AGL studies to learners' subjective confidence level. This connect was in turn considered to be a useful *objective* confidence-based awareness measure that could potentially shed light on the use of implicit and explicit knowledge by adult learners in methodologically similar AGL studies. RT is a standard behavioral measure that can be effortlessly collected in an acceptability judgment task with modern experimental software. We would highly advise future AGL studies to exploit the full potential of RT as a reliable awareness measure to further elucidate the nature of acquired linguistic knowledge.

Acknowledgements

The research was earlier presented in ExLing 2021 and is funded by the Ministry of Science and Technology, Taiwan (107-2410-H-007-002-MY2; 108-2410-H-007-030-MY3). The author is grateful to the following project members for their assistance in the current research (ordered alphabetically by last name): Ssu-Han Chang, Han-Chun Lin, Yi-Shan Lin, Wei-Hsin Lo, Tzu-Hsuan Tseng, and Bo-Ting Yang. The essay has greatly improved with valuable and constructive comments from the editors as well as an anonymous reviewer. As always, the usual disclaimer applies.

References

Bates, Doug & Bolker, Ben & Maechler, Martin & Walker, Steven. 2021. lme4: Linear mixed-effect models using S4 classes. v1.1-28 (http://cran.r-project.org/web/packages/lme4/index.html)

Brysbaert, Marc & Stevens, Michaël. 2018. Power Analysis and Effect Size in Mixed Effects Models: A Tutorial. *Journal of Cognition* 1(1). 9.

Chan, Ricky K.W. & Leung, Janny H.C. 2018. Implicit knowledge of lexical stress rules: Evidence from the combined use of subjective and objective awareness measures. *Applied Psycholinguistics* 39(1). 37–66.

Chan, Ricky K.W. & Leung, Jenny H.C. 2014. Implicit learning of L2 word stress regularities. *Second Language Research* 30(4). 463–484.

Chen, Tsung-Ying. 2010. Some remarks on Contour Tone Units. *Journal of East Asian Linguistics* 19(2). 103–135.

Chen, Tsung-Ying. 2020. An inductive learning bias toward phonetically driven tonal phonotactics. *Language Acquisition*. 27(3). 331–361.

Chen, Tsung-Ying. 2022a. On the learnability of level-based and unit-based tonal OCP generalizations: An artificial grammar learning study. *Glossa: A general journal of linguistics* 7(1). 1–45.

Chen, Tsung-Ying. 2022b. Starting small and starting local: A cognitive source of an inductive bias toward opaque vowel harmony. Hsinchu, Taiwan: National Tsing Hua University. (Manuscript.)

Chen, Tsung-Ying. 2022c. Running online artificial grammar learning experiments with ENIGMA. Hsinchu, Taiwan: National Tsing Hua University. (Manuscript.)

Culbertson, Jennifer. 2012. Typological Universals as Reflections of Biased Learning: Evidence from Artificial Language Learning. *Language and Linguistics Compass* 6(5). 310–329.

Dienes, Zoltán. 2007. Subjective measures of unconscious knowledge. *Progress in Brain Research*. 168. 49–64.

Dienes, Zoltan & Perner, Josef. 1999. A theory of implicit and explicit

knowledge. *Behavioral and Brain Sciences* 22(5). 735–808.
Ellis, Nick C. 2008. Implicit and Explicit Knowledge about Language. In Hornberger, Nancy H (ed.), *Encyclopedia of Language and Education*, 1878–1890. Boston, MA: Springer US.
Gafos, Adamantios I. & Dye, Amanda. 2011. Vowel Harmony: Opaque and Transparent Vowels. In van Oostendorp, Marc & Ewen, Colin J. & Hume, Elizabeth V. & Rice, Keren (eds.), *The Blackwell Companion to Phonology*, 2164–2189. Oxford, UK: John Wiley & Sons, Ltd.
Gandour, Jack. 1981. Perceptual dimensions of tone: Evidence from Cantonese. *Journal of Chinese Linguistics* 9(1). 20–36.
Gandour, Jack. 1983. Tone perception in Far Eastern languages. *Journal of Phonetics* 11(3). 149–175.
Graham, Calbert R. & Williams, John N. 2018. Implicit learning of Latin stress regularities. *Studies in Second Language Acquisition*. 40(1). 3–29.
Granena, Gisela. 2013. Individual Differences in Sequence Learning Ability and Second Language Acquisition in Early Childhood and Adulthood. *Language Learning* 63(4). 665–703.
Finley, Sara. 2015. Learning nonadjacent dependencies in phonology: Transparent vowels in vowel harmony. *Language*. 91(1). 48–72.
Hulstijn, Jan H. 2015. Explaining phenomena of first and second language acquisition with the constructs of implicit and explicit learning: The virtues and pitfalls of a two-system view. In Rebuschat, Patrick (ed.), *Implicit and Explicit Learning of Languages*, 25–46. London, UK: John Benjamins.
Ishikawa, Keiichi. 2019. Incidental and explicit learning of L2 derivational morphology and the nature of acquired knowledge. *Applied Psycholinguistics*. 40(6). 1377–1404.
Lai, Jun & Poletiek, Fenna H. 2013. How "small" is "starting small" for learning hierarchical centre-embedded structures? *Journal of Cognitive Psychology*. 25(4). 423–435.
Leben, William R. 1973. *Suprasegmental phonology*. Cambridge, MA: MIT. (Doctoral dissertation.)
Lichtman, Karen. 2013. Developmental Comparisons of Implicit and Explicit Language Learning. *Language Acquisition*. 20(2). 93–108.
Maie, Ryo & Dekeyser, Robert M. 2020. Conflicting evidence of explicit and implicit knowledge from objective and subjective measures. *Studies in Second Language Acquisition*. 42(2). 359–382.
Martin, Alexander & Peperkamp, Sharon. 2020. Phonetically natural rules benefit from a learning bias: a re-examination of vowel harmony and disharmony. *Phonology*. 37(1). 65–90.
Matuschek, Hannes & Kliegl, Reinhold & Vasishth, Shravan & Baayen,

Harald & Bates, Douglas. 2017. Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*. 94. 305–315.
Moreton, Elliott & Pater, Joe. 2012a. Structure and Substance in Artificial-phonology Learning, Part I: Structure. *Language and Linguistics Compass* 6(11). 686–701.
Moreton, Elliott & Pater, Joe. 2012b. Structure and Substance in Artificial-Phonology Learning, Part II: Substance. *Language and Linguistics Compass* 6(11). 702–718.
Moreton, Elliott & Pertsova, Katya. 2016. Implicit and Explicit Processes in Phonotactic Learning. In Scott, Jennifer & Waughtal, Deb (eds.), *Proceedings of the 40th annual Boston University Conference on Language Development*, 277–290. Somerville, MA: Casadilla Press.
Newport, Elissa L. 1990. Maturational Constraints on Language Learning. *Cognitive Science*. 14(1). 11–28.
Pertsova, Katya & Becker, Misha. 2021. In Support of Phonological Bias in Implicit Learning. *Language Learning and Development* 17(2). 128–157.
R Core Team. 2021. R: a language and environment for statistical computing. v4.1.2. (http://www.r-project.org/)
Reber, Arthur S. 1967. Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior* 6(6). 855–863.
Tsay, Jane S. 2014. A Phonological Corpus of L1 Acquisition of Taiwan Southern Min. In Durand, Jacques & Gut, Ulrike & Kristoffersen, Gjert (eds.), *The Oxford Handbook of Corpus Phonology*, 576–587. Oxford, UK: Oxford University Press.
Wee, Lian-Hee. 2019. *Phonological Tone*. Cambridge, UK: Cambridge University Press.
Zhang, Jie. 2002. *The effects of duration and sonority on contour tone distribution*. New York: Routledge.
Zhang, Jie. 2004. The role of contrast-specific and language-specific phonetics in contour tone distribution. In Hayes, Bruce & Steriade, Donca & Kirchner, Robert (eds.), *Phonetically based Phonology*, 157–190. Cambridge, UK: Cambridge University Press.
Zhang, Jie. 2007. Contour tone distribution is not an artifact of tonal melody mapping. *Studies in the Linguistic Sciences* 33(1). 1–61.
Zhang, Jie & Lai, Yuwen. 2010. Testing the role of phonetic knowledge in Mandarin tone sandhi. *Phonology* 27(1). 153–201.
Zhang, Jie & Lai, Yuwen & Sailor, Craig. 2011. Modeling Taiwanese speakers' knowledge of tone sandhi in reduplication. *Lingua* 121(2). 181–206.
Zhang, Jie & Liu, Jiang. 2016. The productivity of variable disyllabic tone sandhi in Tianjin Chinese. *Journal of East Asian Linguistics* 25(1). 1–35.

Author's address
Tsung-Ying Chen
Department of Foreign Languages and Literature
National Tsing Hua University
No. 101, Sec. 2, Guangfu Road
East District, Hsinchu 300
Taiwan

chen.ty@mx.nthu.edu.tw