Bilingualism: Language and Cognition 16 (4), 2013, 914–923 © Cambridge University Press 2013 doi:10.1017/S136672891300031X

RESEARCH NOTES

Tip-of-the-tongue (TOT) states and cross-linguistic transfer*

KATY BORODKIN
MIRIAM FAUST
Bar-Ilan University

(Received: September 24, 2012; final version received: February 24, 2013; accepted: March 26, 2013; first published online 1 July 2013)

This study examined cross-linguistic transfer in oral language skills in a sample of 50 native Hebrew speakers who learned English as a second language. The ability to retrieve phonological forms of words in naming, as manifested by the tendency to experience tip-of-the-tongue states, was correlated across languages. We also found within and across language correlations between this ability and grammatical accuracy, lexical diversity, and syntactic complexity in second language narratives. These findings are consistent with the transfer across languages in oral language skills and provide insights into the processes linking phonological and higher level encoding in production of connected speech.

Keywords: second language acquisition, oral language production, cross-linguistic transfer, naming, tip-of-the-tongue states

Tip-of-the-tongue (TOT) states are word retrieval failures characterized by a feeling of imminent recall. TOT states can be studied as they naturally occur in everyday life using diary studies (e.g., Burke, MacKay, Worthley & Wade, 1991; Gollan, Montoya & Bonanni, 2005) or in a laboratory setting, where they are evoked by asking participants to name pictures or definitions of low-frequency words (e.g., Brown & McNeill, 1966). The tendency to experience TOT states seems to be sensitive to individual differences, as demonstrated by the findings of more frequent TOT states in individuals with developmental language impairment (Faust, Dimitrovsky & Davidi, 1997), reading-disabled individuals (Faust, Dimitrovsky & Schacht, 2003; Faust & Sharfstein-Friedman, 2003), older adults (Burke et al., 1991), bilinguals (Gollan & Brown, 2006; Gollan et al., 2005), and low-proficiency second language (L2) learners (Borodkin & Faust, 2012, 2013, in press). The current research explored cross-linguistic transfer in the frequency of TOT occurrence. In particular, we examined whether individual differences in the tendency to experience TOT states in native language (L1) naming is related to the corresponding tendency in L2 naming, as well as to L2 connected speech production.

Cross-linguistic transfer has been described as the transfer of knowledge and skills from L1 to L2 (Cummins, 1981), which occurs because both languages are interdependent and rely on a common central processing system (Cummins, 1991). This effect has been observed in bilingual literacy-related skills. For instance, L1 phonological awareness was demonstrated to predict L2 phonological awareness, L2 word decoding (e.g., Durgunoglu, Nagy & Hancin-Bhatt, 1993; Swanson, Rosston, Gerber & Solari, 2008), and L2 reading comprehension (Lindsey, Manis & Bailey, 2003).

The findings supporting the existence of cross-linguistic transfer in oral language skills (spoken and auditory) are far less conclusive and show large variability and inconsistency. For instance, L2 listening comprehension is predictable from L1 listening comprehension, but this correlation varies from $r = -.04$ in some samples (Proctor, August, Carlo & Snow, 2006) to $r = .67$ in others (Abu-Rabia & Siegel, 2002). Similar variability was observed in L1/L2 correlations when oral language skills were assessed in a narrative storytelling task, using measures such as syntactic complexity (Marchman, Martínez-Sussmann & Dale, 2004; Pearson, 2002) and lexical diversity (Kohnert, Kan & Conboy, 2010; Marchman et al., 2004).

A recent meta-analysis of cross-linguistic transfer (Melby-Lervåg & Lervåg, 2011) that included 47 studies yielded moderate-to-high meta-correlations for L1 and L2 literacy-related skills (ranging from $r = .44$ to $r = .60$). In contrast, the L1/L2 meta-correlation in oral language skills was low ($r = .16$). This variance in correlation magnitude has been attributed to differences in the complexity of abilities measured in each domain (Melby-Lervåg & Lervåg, 2011; Proctor et al., 2006). Within the literacy domain, L1 predictor variables are often lower-level aspects of linguistic competence (e.g., L1 phonological awareness and L1 word decoding) that involve the learning of a limited number of sounds and letters, for which a general procedure can be applied in a similar fashion for both languages, thus facilitating cross-linguistic transfer.

* This research was supported in part by grant number 602/06 of the Israel Science Foundation. We wish to thank the three anonymous reviewers for their helpful comments. We also wish to thank the students who participated in the study.

Address for correspondence:
Katy Borodkin, Department of Psychology, Bar-Ilan University, Ramat-Gan, 52900, Israel
Katy.Borodkin@biu.ac.il

On the other hand, L1 predictor variables, routinely examined in the oral language domain (e.g., listening comprehension), are multiply determined and thus less likely to be subjected to such common processes, which makes cross-linguistic transfer harder to detect.

According to the complexity hypothesis (Melby-Lervåg & Lervåg, 2011; Proctor et al., 2006), the lack of strong evidence supporting the existence of cross-linguistic transfer in oral language skills may be attributed to methodological issues rather than to the lack of an actual interdependence between L1 and L2 in this domain. In other words, if skills under investigation are selected with caution, it might be easier to detect cross-linguistic transfer. Elucidating the role of L1 resources available for L2 learners in L2 production may advance our understanding of the individual differences in L2 oral language skills and promote the development of teaching programs that would make better use of these resources in order to improve L2 oral communication skills.

In the current research, we tested the complexity hypothesis by examining the presence of cross-linguistic transfer in a lower-level oral language skill, namely, the ability to retrieve phonological word forms while naming, as reflected in the tendency to experience TOT states. The TOT phenomenon has been used in lexical access models as evidence that words in the mental lexicon are represented at several distinct linguistic levels, such as conceptual-semantic, syntactic, and phonological (Dell, 1986; Levelt, Roelofs & Meyer, 1999). The independence between these levels is apparent in TOT states, which are characterized by the speaker knowing the meaning of the intended word (i.e., it is a kind of vegetable, it is red, and it is round, for the target word *tomato*) and often its syntactic features (e.g., grammatical gender and form class), but being unable to retrieve its sounds (or some of them). Within lexical access models, TOT states are viewed as a success in retrieval of semantic specifications of the word, but a failure to retrieve its phonological information. This retrieval failure might arise when the phonological word forms receive insufficient activation (Burke et al., 1991) or when they are underspecified, fuzzy, or imprecise (Elbro, 1996; Faust et al., 2003).

Whether the locus of TOT states lies within the access to the phonological representations of words or their storage, it is clear that the ability to retrieve phonological word forms is a lower-level oral language skill, as it involves a basic component of speech production system (i.e., single word retrieval) and a specific level of linguistic processing (i.e., phonological). It is certainly not as complex as the skills previously measured in the oral language domain, such as listening comprehension, which involves numerous levels of linguistic processing and several components of the speech perception system, including the perceptual analysis of stress and rhythm patterns, tone patterns and intonational contours, vocabulary knowledge, and the analysis of the syntactic constituents, such as the subject, verb, and object of a sentence (J. C. Richards, 1983). Since the retrieval of phonological word forms while engaged in L1 naming resembles more the predictor variables assessed in the literacy domain than the oral language domain in its relative lack of complexity, we expected that this ability would be correlated across languages.

The complexity hypothesis (Melby-Lervåg & Lervåg, 2011; Proctor et al., 2006) suggests that cross-linguistic transfer can be evident not only in the prediction of lower-level L2 skills, but also in more complex and multifaceted L2 skills, as long as the L1 predictor variables are sufficiently basic. This is supported by findings in literacy domain showing that L2 reading comprehension is related to L1 phonological awareness and L1 word decoding (Lindsey et al., 2003; Melby-Lervåg & Lervåg, 2011). Given this evidence, we expected that the tendency to experience TOT states in L1 naming would be associated not only with the corresponding tendency in L2 naming, but also with the more complex and multiply determined aspects of L2 oral language production, such as production of connected speech. In the current research, L2 connected speech production was evaluated in a storytelling task, using three measures: grammatical accuracy, which is the ability to produce an error-free linguistic output (Lennon, 1990); syntactic complexity, which refers to the range of forms in language production and the level of sophistication in these forms (Ortega, 2003); and lexical diversity, which reflects the range of vocabulary displayed in the output (Durán, Malvern, Richards & Chipere, 2004).

The suggested relationship between TOT states and continuous language production is consistent with the predictions of sentence production models (Bock & Levelt, 1994; Dell, 1986). These models agree that sentence construction, like single word retrieval, occurs at several hierarchically ordered levels of representation, starting from conceptual-semantic, syntactic, morphological, and phonological encoding, and ending with motor encoding. Although they differ in the degree of interactivity assumed among the processing levels, both serial (Bock & Levelt, 1994) and interactive (Dell, 1986) models predict that phonological encoding might affect higher level encoding. Such effects are attributed to a self-monitoring mechanism within the serial models and to bottom–up activation within the interactive models. There is some experimental evidence, though surprisingly scarce, supporting these predictions (Bock, 1987; Jaeger, Furth & Hilliard, 2012; Lee & Gibbons, 2007). In terms of sentence production models, TOT states, which interfere with phonological encoding of the planned items, may negatively affect the production of continuous language, as measured by grammatical accuracy, lexical diversity, and syntactic complexity, if

the latter measures are regarded as indices of higher level encoding (i.e., morphological, syntactic, and conceptual-semantic).

Sentence production models (Bock & Levelt, 1994; Dell, 1986) predict WITHIN LANGUAGE negative correlations between the tendency to experience TOT states and connected speech production. Accordingly, we expected that greater tendency for L2 TOT states would be related to reduced grammatical accuracy, lexical diversity and syntactic complexity in L2 narratives. More importantly to this study, which focused on cross-linguistic transfer in the oral language domain, we also expected to find evidence of correlations ACROSS LANGUAGES. In line with the complexity hypothesis (Melby-Lervåg & Lervåg, 2011; Proctor et al., 2006), a positive L1/L2 correlation in the tendency to experience TOT states and negative correlations between L1 TOT states and measures of L2 continuous language production were predicted.

Methods

Participants

The sample included 50 (34 females) undergraduate students at Bar-Ilan University, aged 22.63 years on average ($SD = 2.24$), who were recruited through advertisements around campus. Students enrolled in an introductory psychology course received academic credit for participation. Others were paid $40 to $80, depending on the time they spent in the lab. Participants were native Hebrew speakers, who were not familiar with any language other than Hebrew and English. They have learned English as L2 in a formal school setting for about 11 years, starting at age 7, and none had an immersion experience of more than 3 months in an English-speaking country.

Materials and procedures

Participants were administered the tasks described here as part of a larger battery (Borodkin & Faust, 2013, in press), to complete which they spent between 3.5 and 7.5 hours in the lab. Participants were tested in two or three separate sessions, with each session lasting up to 3 hours. A trained experimenter administered the tasks described here individually, except for Raven's Standard Progressive Matrices (Raven, Raven & Court, 1993; see below), which was given either individually or in pairs. The order of tasks administration was randomized.

Tip-of-the-tongue experiments

The participants completed a TOT naming experiment in each language. The target words used in the L1 and L2 experiments were different. The experimental stimuli were black-and-white photographs of objects and living things, which were collected through Internet search. The stimuli for each experiment were selected on the basis of the results of a pilot study that included 15 students who did not take any further part in the research. They were presented with 151 pictures to be named in Hebrew and 160 pictures to be named in English. Following the pilot study, 12 stimuli were excluded from the final list designated for the L1 experiment, because the participants were not visually familiar with the depicted object, although they knew the word (i.e., they knew what *anvil* means, but not what the object looks like). Eight stimuli to be named in English were omitted due to low familiarity with the target word (e.g., *faucet*). The final list of stimuli presented in the L1 and L2 experiments included 139 and 152 photographs, respectively.

The selection of target words was aimed at accounting for the more limited vocabulary knowledge and language use in L2 compared to L1. Consequently, the L1 list contained low-frequency words such as *abacus* and *tripod* (the full list of stimuli is available in Borodkin & Faust, 2013), which might have been too difficult to name if presented in the L2 experiment and thus might have elicited too many "Don't know" responses. The L2 list, on the other hand, consisted of high-frequency words, such as *knife* and *stairs* (for a full list of stimuli, see Borodkin & Faust, in press), which might have been too easy if used in the L1 experiment and thus might have evoked too few TOT responses. Given (to the best of our knowledge) the lack of a Hebrew database that provides word frequency values directly comparable to those reported in English (e.g., Brown Corpus; Kucera & Francis, 1967), we confirmed the differences between L1 and L2 word lists using subjective frequency ratings. Ten participants who did not take any further part in this research were asked to rate how frequent the words in each list were in the respective language (on a scale of 1 = not frequent at all to 10 = very frequent). A *t*-test for independent samples yielded a significant result, $t(289) = 7.14$, $p < .001$, indicating that stimuli in L1 ($M = 4.77$, $SD = 1.62$) were less frequent than in L2 ($M = 5.92$, $SD = 1.09$). The lists also differed in word length (assessed by the number of letters), $t(289) = 3.10$, $p = .002$, such that L1 words ($M = 5.14$, $SD = 1.46$) were shorter than L2 words ($M = 5.74$, $SD = 1.84$).

The experimental procedures and the scoring criteria were the same for the L1 and L2 experiments. Both experiments were administered using the E-prime (version 1.2) software (Schneider, Eschman & Zuccolotto, 2002), which was controlled by the experimenter according to participants' responses. Pictures were presented in a random order with no time limit. Responses were collected by the program, and in some cases the records were completed manually by the experimenter. Each experiment lasted from 30 to 90 minutes.

Following the presentation of a target picture, participants could either correctly name it, provide a

https://doi.org/10.1017/S136672891300031X Published online by Cambridge University Press

"Don't know" response, or report being in a TOT state, which they were told occurs when the speaker knows the word, but cannot say it at that moment. Except for pictures that triggered correct responses, all trials ended with a recognition procedure, in which participants were presented with the target word and asked if they were familiar with it. On TOT trials, participants were also asked if the target word matched the word they were searching for. An initial TOT response was scored as a positive TOT if upon the presentation of the target word participants recognized it and confirmed that it was the word they were searching for; a negative TOT was coded when the participants recognized the target word but searched for another word. When participants were unfamiliar with the target word, the initial TOT response was scored as a "Don't know" trial.

Among the various types of naming responses in a TOT experiment, positive TOT states are the most straightforward measure of the ability to retrieve phonological word forms, which was of interest here. Thus, for the purposes of the current study, only positive TOT responses were analyzed (whenever TOT states are mentioned throughout the manuscript, this type is intended). The analyses were conducted on raw and relative TOT scores, which were calculated, following Gollan and Brown (2006), as the number of TOT responses divided by the sum of correct and TOT responses. Relative scores account for the number of opportunities to experience TOT states, and thus provide a way to control for differences in vocabulary knowledge (Gollan & Brown, 2006), which may be large in L2 learners.

Connected speech production in second language

The production of continuous language in L2 was assessed using a storytelling task, which consisted of two separate sets of sequenced pictures adapted from children books. Story 1 (Kulot, 2003) included 15 pictures describing a giraffe and a crocodile that fell in love and wanted to move in together. At first, they tried the crocodile's house, and when this did not work out, they moved on to the giraffe's house. In the end they built a house that was suited for them both. Story 2 (Donaldson, 2001) included 14 pictures that depicted a witch who traveled riding her broom. Every now and then she had lost something, and while searching for it, she had encountered new friends, who had joined her on her journey. Towards the end of the story, the witch was attacked by a dragon and had been saved by her friends.

The participants were asked to tell each story twice, first in Hebrew and then in English. This was done to reduce inter-individual variation in the interpretation of the story plot. To ensure that the narratives were of sufficient length, participants were instructed to provide approximately three sentences for each picture. Participants required approximately 10 minutes to complete each story and were recorded while narrating in English.

The narratives were evaluated for grammatical accuracy, syntactic complexity, and lexical diversity. The scores on each of the measures were averaged across the two stories. To calculate grammatical accuracy and syntactic complexity, the data were first segmented into analysis of speech units (AS-units). An AS-unit was previously defined as "a single speaker's utterance consisting of an independent clause, or sub-clausal unit, together with any subordinate clause(s) associated with either" (Foster, Tonkyn & Wigglesworth, 2000, p. 365). The AS-unit was chosen as the basic productive unit because it reflects more reliably syntactic complexity as compared to sentence-based or clause-based units (Foster et al., 2000).

Grammatical accuracy was calculated as the ratio of error-free AS-units relative to the total number of AS-units. The equivalent measure in the written data (i.e., the mean number of error-free T-units, the latter being a minimal terminal unit) was found to be one of the most valid measures of accuracy (Wolfe-Quintero, Inagaki & Kim, 1998). Grammatical accuracy was scored according to the guidelines proposed by Polio (1997), who described six error categories. These include sentence structure errors (e.g., word order errors, such as *I can't tell the all story*), verb-centered errors (e.g., subject–verb agreement errors, such as *They all was so happy*), reference errors (e.g., quantifier–noun agreement errors, such as *In this days*), word-level choice (e.g., incorrect use of idiomatic phrases, as in *on the second side* instead of *on the other hand*), article errors (e.g., omission of genitives, as in *So they try to live in the crocodile house*), and punctuation errors. An error-free AS-unit did not contain any of the errors suggested by Polio, except for errors irrelevant in oral language production (e.g., punctuation errors).

Syntactic complexity was evaluated using the mean length of unit (MLU), calculated here as the total number of words produced during the narration divided by the total number of AS-units. The MLU is one of the most common measures employed to estimate syntactic complexity; it has adequate concurrent (Wolfe-Quintero et al., 1998) and predictive validity (Ortega, 2003).

Lexical diversity was evaluated by the D measure, which was developed by Malvern and Richards (2002) to overcome problems associated with calculation of the simple type–token ratio. The type–token ratio is formed from the division of the number of types (different words) by the total number of words (tokens). Unfortunately, it varies as a function of sample length, such that longer samples tend to produce lower ratio values (B. Richards, 1987). The D measure is based on the analysis of the probability for new words to appear in longer samples of speech or writing. This yields a mathematical model

Table 1. *Means (M), standard deviations (SD) and intercorrelations for L1/L2 TOT states, L2 oral narrative measures, and nonverbal intelligence.*

Measure	*M*	*SD*	Possible range	1.	2.	3.	4.	5.	6.	7.
1. L1 raw TOT scores	12.32	6.24	1–33	–						
2. L2 raw TOT scores	19.70	13.54	2–76	.31*	–					
3. L1 relative TOT scores	0.10	0.05	0.01–0.28	.99***	.30*	–				
4. L2 relative TOT scores	0.21	0.15	0.01–0.64	.39**	.85***	.41**	–			
5. D measure	35.73	7.97	17.22–53.93	–.27†	–.30*	–.29*	–.45**			
6. Grammatical accuracy	0.61	0.20	0.21–0.92	–.32*	–.39**	–.33*	–.63***	.66***	–	
7. MLU	9.03	1.00	6.63–10.76	–.30*	–.33*	–.32*	–.38**	.16	.04	–
8. Raven's Matrices	52.44	4.46	40–60	–.20	–.03	–.22	–.02	.15	.10	–.03

L1 = native language, L2 = second language, MLU = mean length of unit, TOT = tip-of-the-tongue
†$p = .054$, *$p < .05$, **$p < .01$, ***$p < .001$

of how type–token ratio varies with token size. The D measure is obtained by comparing the mathematical model with the empirical data. Higher D values are considered to reflect higher lexical diversity. In order to obtain the D measure, the data were transferred to CHAT format and analyzed using the *vocd* program (MacWhinney, 2000) included in the CLAN software (MacWhinney, 1995) of the CHILDES project.

For 20% of the data, scores involving qualitative analysis (i.e., AS-units and grammatical accuracy) were coded by a second rater. The inter-rater reliability, computed using intraclass correlation coefficient, was sufficiently high for both AS-units and grammatical accuracy (.95 and .85, respectively, *p*s < .001).

Nonverbal intelligence

We tested the assumption that TOT responses are a specific measure of the ability to retrieve phonological word forms while naming rather than a measure of an overall intellectual ability. To this end, we evaluated the association between L1/L2 TOT states and performance on a nonverbal intelligence test, namely, Raven's Standard Progressive Matrices (Raven et al., 1993). This test involves complex visual patterns with a missing part. Below each of the patterns, six or eight segments are presented, one of which correctly completes the larger pattern. The test was administered with a 20-minute time limit. The number of correct answers was counted.

Results

Within and across language correlations

As demonstrated in Table 1, the ability to retrieve phonological word forms in naming was positively and significantly correlated across languages, in both raw and relative scores. Consistent with the cross-linguistic hypothesis, individuals with frequent TOT states in L1 naming were also more likely to experience frequent TOT states in L2 naming. Furthermore, the measures of L2 connected speech were negatively and significantly correlated within and across language with the tendency to experience TOT states, as measured in both raw and relative scores. Individuals with greater tendency for TOT states in L1/L2 naming displayed lower grammatical accuracy, syntactic complexity, and lexical diversity in L2 narratives. The linguistic meaning of these correlations is illustrated in Supplementary Materials accompanying the online version of this paper (see the journal's webpage at http:// journals.cambridge.org/BIL), which provide actual narratives from two participants – one with few and the other with many L1/L2 TOT states.

Although the correlations between TOTs and measures of L2 connected speech production showed similar trends, regardless of TOT calculation method, they tended to be greater using relative compared to raw scores, especially for L2 TOTs (for example, the correlation with grammatical accuracy increased from $r = –.39$ to $r = –.63$ when calculated on L2 raw compared to relative TOT scores, respectively). As detailed in the Discussion below, we attribute this difference to raw TOT responses being a less reliable measure of phonological word form retrieval, particularly in L2 naming. Thus, further analyses of the data were conducted using relative scores only.

Unique contributions in prediction of connected speech production

As both across and within language correlations were significant, we also conducted multiple regression analyses predicting each of the measures of L2 continuous language production from L1 and L2 scores simultaneously. We reasoned that if both kinds of correlations mainly represent the effects of phonological

https://doi.org/10.1017/S136672891300031X Published online by Cambridge University Press

encoding on higher level encoding in sentence production, than L1 and L2 TOTs will explain largely overlapping portions of variability in L2 connected speech measures. Entering L1 and L2 scores simultaneously as predictors will thus result in nonsignificant unique contributions. If, on the other hand, the correlations are affected by other factors in addition to phonological encoding (e.g., L2 proficiency might play a role in the within language correlations), than L2 and/or L1 TOT states might have significant unique contributions to variability in L2 lexical diversity, grammatical accuracy, and syntactic complexity.

For grammatical accuracy, the overall model was significant, $F(2,47) = 16.17$, $p < .001$, $R^2 = .41$. Individuals with greater tendency for TOT states while naming in L2 had lower grammatical accuracy in L2 narratives, $\beta = -.60$, $t(47) = 4.86$, $p < .001$. The across language association between L1 TOTs and L2 grammatical accuracy was reduced and no longer statistically significant, when L2 TOT states were entered into the equation, $\beta = -.09$, $t(47) = 0.72$, $p = .47$. The multiple regression analysis on MLU scores also yielded a significant result, $F(2,47) = 5.07$, $p = .01$, $R^2 = .18$. Individuals with greater tendency for L2 TOT states were likely to provide shorter AS-units, $\beta = -.30$, $t(47) = 2.09$, $p = .04$. After entering L2 TOT states into the equation, the across language association between L1 TOT states and MLU in L2 narratives was decreased and no longer statistically significant, $\beta = -.20$, $t(47) = 1.35$, $p = .19$. The analysis also yielded a significant result for the D measure, $F(2,47) = 6.40$, $p = .003$, $R^2 = .21$. Individuals who experienced more TOT states in L2 naming had lower D scores in L2 narratives, $\beta = -.39$, $t(47) = 2.76$, $p = .008$. After accounting for L2 TOT states, the association between L1 TOT states and the D measure was no longer statistically significant, $\beta = -.14$, $t(47) = 0.96$, $p = .34$. Thus, regression analyses revealed that within language TOT rates contributed more to the variability in L2 connected speech than across language TOT rates.

Matching L1 and L2 lists for word frequency and length

As described above, L1 words were, on average, less frequent and shorter compared to L2 words. To evaluate the effects these differences might have had on the results described above, we re-ran the analyses after matching the lists on item difficulty. This was done by omitting the least frequent and the shortest 24 words from the L1 list and the most frequent and the longest 39 words from the L2 list. The remaining 115 L1 words and 113 L2 words did not differ significantly in frequency, $t(226) = 1.89$, $p = .06$, ($M = 5.18$, $SD = 1.43$ and $M = 5.48$, $SD = 0.88$, respectively) or length, $t(226) = 1.63$, $p = .11$ ($M = 5.30$, $SD = 1.45$ and $M = 5.66$, $SD = 1.85$, respectively). Relative TOT scores were re-calculated including only the stimuli in the matched lists and then used to re-examine the effects of cross-linguistic transfer.

The results obtained using matched and complete word lists were very similar. The tendency to experience TOT states was significantly correlated across languages, $r = .44$, $p = .001$. The across language correlations between TOT states and measures of L2 connected speech ranged from $r = -.22$ to $r = -.33$, and the corresponding within language correlations ranged from $r = -.34$ to $r = -.63$. Multiple regression analyses predicting each of the measures of L2 connected speech, based on L1 and L2 TOT scores, yielded a significant unique contribution to L2 TOT states and a non-significant contribution to L1 TOT states. These results indicate that the differences between L1 and L2 word lists in item difficulty did not affect the results observed in the analyses using complete word lists.

Tip-of-the-tongue states and nonverbal intelligence

Lastly, we calculated the correlations between L1/L2 TOT states and performance on Raven's Matrices, which, as Table 1 above demonstrates, were low and non-significant. Thus, we ruled out the possibility that TOT states reflect a general intellectual ability rather than a specific ability to retrieve phonological word forms.

Discussion

According to the complexity hypothesis (Melby-Lervåg & Lervåg, 2011; Proctor et al., 2006), cross-linguistic transfer in oral language skills has been difficult to identify in previous studies due to the complexity of the predictor variables applied in this domain. This hypothesis was tested in this study, which demonstrated that the ability to retrieve phonological word forms in L1 naming was related to the corresponding ability in L2 naming as well as to measures of L2 continuous language production. Individuals experiencing frequent TOT states in L1 naming were also more likely to report frequent TOT states in L2 naming and exhibited diminished grammatical accuracy, syntactic complexity, and lexical diversity while narrating in L2. Assuming that the tendency to experience TOT states is a lower-level oral language skill, which was validated by the lack of correlation between this measure and a measure of nonverbal intelligence (Raven's Matrices; Raven et al., 1993), our findings provide evidence for cross-linguistic transfer in oral language skills and support the complexity hypothesis.

The findings of this study also contribute to the under-researched topic of the phonological effects on higher level encoding in sentence production (Bock & Levelt, 1994; Dell, 1986). In one of the few studies that explored such effects, inaccessibility of phonological

https://doi.org/10.1017/S136672891300031X Published online by Cambridge University Press

word forms (elicited by phonological priming) was demonstrated to prompt a revision of word order in sentences (Bock, 1987). In another study (Lee & Gibbons, 2007), syntactic choice of inclusion/omission of the complementizer *that* in sentences like *Henry knew* (*that*) *Lucy*/*Louise washed the dishes* was found to be related to phonological preference of alternating sequence of stressed and unstressed syllables. Lastly, it was reported (Jaeger et al., 2012) that lexical selection of a verb form (e.g., *handed*/*passed*) depended on whether the onset of the verb and the subject of the sentence phonologically overlap (e.g., *Hannah handed the flask to the child* vs. *Hannah passed the flask to the child*).

The relationship between continuous language production measures and TOT naming failures might be regarded as further evidence of phonological processing effects on higher level encoding in sentence production. Frequent TOT states, which are likely to be a manifestation of decreased availability of phonological word forms either due to access (Burke et al., 1991) or storage problems (Elbro, 1996; Faust et al., 2003), might interfere with successful completion of phonological encoding of the lexical entries chosen for production. This, in turn, might negatively affect morphological, syntactic, and conceptual-semantic encoding of a sentence. The negative correlations found between the tendency to experience TOT states in L1/L2 naming and grammatical accuracy, lexical diversity, and syntactic complexity in L2 narratives might be a reflection of such processes, if the latter measures are treated as indices of higher level encoding during sentence production. This interpretation, however, would have been more powerful, if we had been able to demonstrate L1–L1 correlations between measures of connected speech production and TOT states.

Additional factors may play a role contributing to the association between the ability to retrieve phonological word forms and connected speech production, as indicated by differences in the strength of within compared to across language correlations. L2 (relative) TOT scores seemed to be more strongly correlated with L2 grammatical accuracy, lexical diversity, and syntactic complexity compared to the corresponding L1 scores (see Table 1 above). Multiple regression analyses confirmed that TOT scores had a unique within language, but not across language, contribution to variability in the measures of L2 continuous language production. These differences remained unchanged even after partialling out L1/L2 differences in item difficulty, as was demonstrated in the analyses using stimuli lists matched on word frequency and length.

One factor that might have determined the stronger within language association was L2 proficiency. The three measures of continuous language production applied in this study, namely, grammatical accuracy, syntactic complexity, and lexical diversity, are traditionally used as measures of L2 proficiency. Furthermore, L2 proficiency has been previously observed to affect lexical access (Kroll, Michael, Tokowicz & Dufour, 2002). Assuming that TOT states are a measure of L2 lexical access, the stronger within than across language correlations (and the consequent greater unique contributions observed in multiple regression analyses) between TOT states and L2 continuous language production might have resulted, at least in part, from the relationship between L2 proficiency and L2 lexical access. This account, however, remains to be corroborated in future research using additional L2 proficiency measures. Although L2 grammatical accuracy, lexical diversity, and syntactic complexity have been validated against other L2 proficiency measures, such as holistic ratings and program levels (Iwashita, Brown, McNamara & O'Hagan, 2008; Ortega, 2003), they were found to depend on factors other than L2 proficiency as well (Ortega, 2003; Tavakoli & Foster, 2008). A worthy alternative to these measures is the profiling procedures suggested by Processability Theory (Pienemann, 1998).

In the context of measures of L2 continuous language production, the differential correlations between these measures and L1/L2 TOT states should be also noted. Specifically, L1/L2 (relative) TOT scores seemed to be correlated more strongly with grammatical accuracy than with lexical diversity or syntactic complexity (see Table 1). This difference might have resulted from the characteristics of the narrative task applied in the current study, which was based on children books. The plot in both stories was highly repetitive (e.g., in Story 2 the witch loses an item once in a while, and each time that this occurs, she lands on the ground, searches for it, and meets a new friend). Such repetitiveness might have affected measures of linguistic variability (i.e., syntactic complexity and lexical diversity) to a greater extent compared to grammatical accuracy, which might have led to less variability in these measures and consequently–to reduced correlations between these measures and TOT rates.

Our last remark concerns the effects of TOT calculation method on the strength of correlations between TOT scores and measures of L2 continuous language production. For L1 TOTs, these correlations remained largely unaltered regardless of whether calculated on raw or relative scores. For L2 TOTs, on the other hand, the correlations were higher when computed using relative compared to raw scores (see Table 1). To account for these differential effects, we considered TOT states in relation to correct responses in each language, which is illustrated using raw scores in Figure 1. As can be seen, there was a fairly consistent trend ($r = -.84$) in L1 naming responses, such that participants who provided many correct responses also had few TOT responses and vice versa. In other words, having many opportunities to experience TOT states often resulted in few TOT naming

https://doi.org/10.1017/S136672891300031X Published online by Cambridge University Press

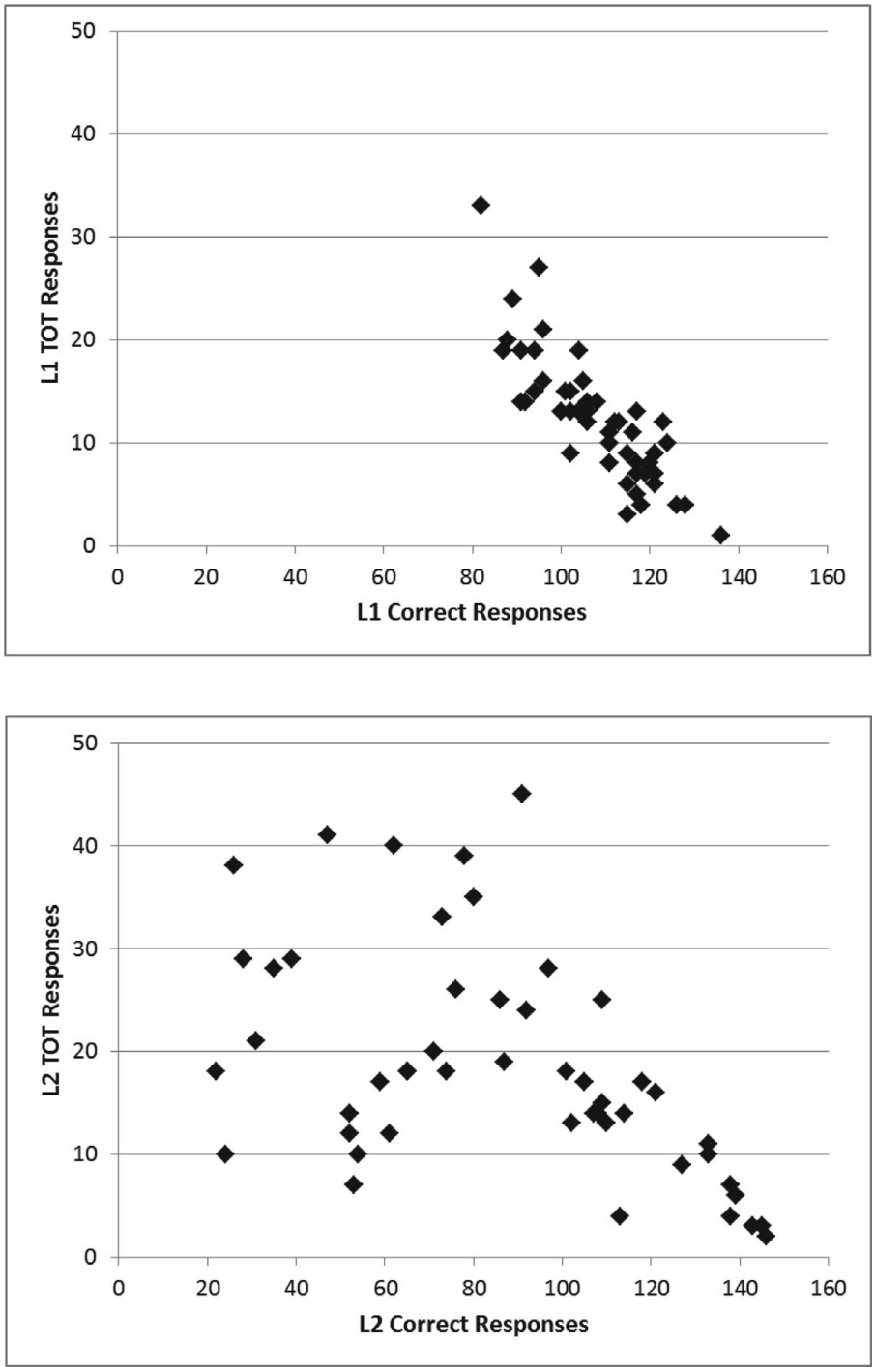

Figure 1. The association between TOT states and correct responses (both in raw numbers) in native language (L1; upper panel) and second language (L2; lower panel) naming.

failures; therefore, raw and relative TOT scores provided a similar estimation of the ability to retrieve phonological word forms in L1 naming.

In L2 naming, many correct responses were also associated with few TOT naming failures, but this trend was not as consistent as in L1 naming ($r = -.51$). As Figure 1 demonstrates, there was a group of low L2 proficiency individuals who named correctly very few L2 target words (20–60 out of 152) and reported relatively few L2 TOT states (10–20). Thus, naming in L2 – but not in L1 – in some cases elicited few TOT states, but also provided few opportunities for TOT states. In such individuals, raw TOT scores overestimated the ability to retrieve phonological word forms in L2 naming, which, as

https://doi.org/10.1017/S136672891300031X Published online by Cambridge University Press

a result, led to decreased correlations between these scores and measures of L2 continuous language production. Thus, the results of this study suggest that whereas L1 naming responses are not particularly affected by the method of TOT calculation, TOT responses in L2 naming are more reliably indexed by relative than raw scores. This difference is related to greater individual variability in L2 compared to L1 vocabulary knowledge, as indirectly indicated by the number of correct responses in L1 and L2 TOT experiments.

To summarize, the present study provides evidence that the ability to retrieve phonological word forms is transferred across languages. We also demonstrated that this lower-level phonological skill can predict the more complex aspects of L2 continuous language production. These findings have a theoretical as well as practical value, as they might be helpful in developing more sophisticated L2 instruction programs, especially for individuals with difficulties in L2 language production, which aim to improve the ability to retrieve phonological word forms in L1 and L2.

References

Abu-Rabia, S., & Siegel, L. S. (2002). Reading, syntactic, orthographic, and working memory skills of bilingual Arabic–English speaking Canadian children. *Journal of Psycholinguistic Research, 31,* 661–678.

Bock, K. (1987). An effect of the accessibility of word forms on sentence structures. *Journal of Memory and Language, 26,* 119–137.

Bock, K., & Levelt, W. J. (1994). Language production: Grammatical encoding. In M. A. Gernsbacher (ed.), *Handbook of psycholinguistics*, pp. 945–984. London: Academic Press.

Borodkin, K., & Faust, M. (2012). Word retrieval in developmental language impairments: Application of the Tip-of-the-Tongue paradigm. In M. Faust (ed.), *The handbook of the neuropsychology of language*, pp. 963–982. Oxford: Wiley-Blackwell.

Borodkin, K., & Faust, M. (2013). Native language phonological skills in low-proficiency second language learners. Ms., Bar-Ilan University.

Borodkin, K., & Faust, M. (in press). Naming abilities in low-proficiency second language learners. *Journal of Learning Disabilities,* doi:10.1177/0022219412453769. Published online by Sage, August 27, 2012.

Brown, R., & McNeill, D. (1966). The "tip of the tongue" phenomenon. *Journal of Verbal Learning and Verbal Behavior, 5,* 325–337.

Burke, D. M., MacKay, D. G., Worthley, J. S., & Wade, E. (1991). On the tip of the tongue: What causes word finding failures in young and older adults? *Journal of Memory and Language, 30,* 542–579.

Cummins, J. (1981). The role of primary language development in promoting educational success for language minority students. In California State Department of Education (ed.), *Schooling and language minority students: A theoretical framework*, pp. 3–49. Los Angeles, CA: Evaluation, Dissemination, and Assessment Center, California State University.

Cummins, J. (1991). Interdependence of first and second-language proficiency in bilingual children. In E. Bialystok (ed.), *Language processing in bilingual children*, pp. 70–89. Cambridge: Cambridge University Press.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 93,* 283–321.

Donaldson, J. (2001). *Room on the broom*. London: Macmillan Children's Books.

Durán, P., Malvern, D., Richards, B., & Chipere, N. (2004). Developmental trends in lexical diversity. *Applied Linguistics, 25,* 220–242.

Durgunoglu, A. Y., Nagy, W. E., & Hancin-Bhatt, B. J. (1993). Cross-language transfer of phonological awareness. *Journal of Educational Psychology, 85,* 453–465.

Elbro, C. (1996). Early linguistic abilities and reading development: A review and a hypothesis. *Reading and Writing, 8,* 453–485.

Faust, M., Dimitrovsky, L., & Davidi, S. (1997). Naming difficulties in language-disabled children: Preliminary findings with the application of the Tip-of-the-Tongue paradigm. *Journal of Speech, Language and Hearing Research, 40,* 1026–1036.

Faust, M., Dimitrovsky, L., & Schacht, T. (2003). Naming difficulties in children with dyslexia: Application of the tip of the tongue paradigm. *Journal of Learning Disabilities, 36,* 203–216.

Faust, M., & Sharfstein-Friedman, S. (2003). Naming difficulties in adolescents with dyslexia: Application of the tip-of-the-tongue paradigm. *Brain and Cognition, 53,* 211–217.

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics, 21,* 354–375.

Gollan, T. H., & Brown, A. S. (2006). From tip-of-the-tongue (TOT) data to theoretical implications in two steps: When more TOTs means better retrieval. *Journal of Experimental Psychology: General, 135,* 462–483.

Gollan, T. H., Montoya, R. I., & Bonanni, M. P. (2005). Proper names get stuck on bilingual and monolingual speakers' tip of the tongue equally often. *Neuropsychology, 19,* 278–287.

Iwashita, N., Brown, A., McNamara, T., & O'Hagan, S. (2008). Assessed levels of second language speaking proficiency: How distinct? *Applied Linguistics, 29,* 24–49.

Jaeger, T. F., Furth, K., & Hilliard, C. (2012). Phonological overlap affects lexical selection during sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38,* 1439–1449.

Kohnert, K., Kan, P. F., & Conboy, B. T. (2010). Lexical and grammatical associations in sequential bilingual preschoolers. *Journal of Speech, Language and Hearing Research, 53,* 684–698.

Kroll, J. F., Michael, E., Tokowicz, N., & Dufour, R. (2002). The development of lexical fluency in a second language. *Second Language Research, 18,* 137–171.

Kucera, H., & Francis, W. N. (1967). *Computational analysis of present-day American Engish*. Providence, RI: Brown Unviersity Press.

Kulot, D. (2003). *Das kleine krokodil und die große liebe*. Stuttgart-Wien: K. Thienemann Verlag (Thienemann Verlag GmbH).

Lee, M.-W., & Gibbons, J. (2007). Rhythmic alternation and the optional complementiser in English: New evidence of phonological influence on grammatical encoding. *Cognition, 105,* 446–456.

Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning, 40,* 387–417.

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22,* 1–38.

Lindsey, K. A., Manis, F. R., & Bailey, C. E. (2003). Prediction of first-grade reading in Spanish-speaking English-language learners. *Journal of Educational Psychology, 95,* 482–494.

MacWhinney, B. (1995). *The CHILDES project: Tools for analyzing talk* (2nd edn.). Hillsdale, NJ: Lawrence Erlbaum.

MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk* (3rd edn.). Mahwah, NJ: Lawrence Erlbaum.

Malvern, D., & Richards, B. (2002). Investigating accommodation in language proficiency interviews using a new measure of lexical diversity. *Language Testing, 19,* 85–104.

Marchman, V. A., Martínez-Sussmann, C., & Dale, P. S. (2004). The language-specific nature of grammatical development: Evidence from bilingual language learners. *Developmental Science, 7,* 212–224.

Melby-Lervåg, M., & Lervåg, A. (2011). Cross-linguistic transfer of oral language, decoding, phonological awareness and reading comprehension: A meta-analysis of the correlational evidence. *Journal of Research in Reading, 34,* 114–135.

Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics, 24,* 492–518.

Pearson, B. (2002). Narrative competence among monolingual and bilingual school children in Miami. In K. Oller & R. Eilers (eds.), *Language and literacy in bilingual children*, pp. 135–174. Clevedon: Multilingual Matters.

Pienemann, M. (1998). *Language processing and second language development: Processability Theory*. Amsterdam: John Benjamins.

Polio, C. G. (1997). Measures of linguistic accuracy in second language writing research. *Language Learning, 47,* 101–143.

Proctor, C. P., August, D., Carlo, M. S., & Snow, C. (2006). The intriguing role of Spanish language vocabulary knowledge in predicting English reading comprehension. *Journal of Educational Psychology, 98,* 159–169.

Raven, J., Raven, J. C., & Court, J. H. (1993). *Manual for Raven's Progressive Matrices and Vocabulary Scales*. Oxford: Oxford Psychologists Press.

Richards, B. (1987). Type/token ratios: What do they really tell us? *Journal of Child Language, 14,* 201–209.

Richards, J. C. (1983). Listening comprehension: Approach, design, procedure. *TESOL Quarterly, 17,* 219–240.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime User's Guide (Version 1.2)*. Pittsburgh, PA: Psychology Software Tools Inc.

Swanson, H. L., Rosston, K., Gerber, M., & Solari, E. (2008). Influence of oral language and phonological awareness on children's bilingual reading. *Journal of School Psychology, 46,* 413–429.

Tavakoli, P., & Foster, P. (2008). Task design and second language performance: The effect of narrative type on learner output. *Language Learning, 58,* 439–473.

Wolfe-Quintero, K., Inagaki, S., & Kim, H. Y. (1998). *Second language development in writing: Measures of fluency, accuracy and complexity*. Honolulu, HI: Second Language Teaching and Curriculum Center, University of Hawai'i at Mānoa.