
AI translation · View original & related papers at
chinaxiv.org/items/chinaxiv-201812.00203

Cognitive Aging of the Tip-of-the-Tongue Phenomenon in Naturalistic Settings: A Diary Study

Authors: Zhao Ruiying, Lou Hao, Ouyang Mingkun, Zhang Qingfang, Zhang Qingfang

Date: 2018-12-20T00:00:00+00:00

Abstract

The tip-of-the-tongue (TOT) effect is a word retrieval failure phenomenon where a word is on the tip of one's tongue but cannot be articulated. This study conducted a 28-day longitudinal tracking of young and older adults using TOT retrospective questionnaires and diary recording methods to investigate the occurrence characteristics and mechanisms of cognitive aging of the tip-of-the-tongue effect in naturalistic settings. The results showed: (1) In naturalistic settings, TOT exhibits a cognitive aging phenomenon, with older adults showing higher TOT occurrence frequency than young adults; (2) When TOT occurs, individuals retrieve alternative words and semantically related information, and young adults experience stronger metacognitive experiences of excitement and fatigue than older adults; (3) After TOT occurs, target words can almost always be retrieved, with older adults requiring longer TOT resolution time than young adults, but with no age differences in resolution rate. TOT resolution time is influenced by cognitive and metacognitive factors such as target word familiarity, presence of alternative words, and individual subjective experience. (4) Recording the cognitive and metacognitive states of TOT affects the subsequent occurrence state of TOT and prolongs its resolution time, indicating that participants expended more effort and time to resolve TOT. Both cognitive and metacognitive state factors of TOT exerted an influence on TOT resolution time.

Full Text

Aging of the Tip-of-the-Tongue in Natural Contexts: A Diary Study

ZHAO Ruiying, LOU Hao, OUYANG Mingkun, ZHANG Qingfang
Department of Psychology, Renmin University of China, Beijing 100872, China

Abstract

The tip-of-the-tongue (TOT) phenomenon refers to a word retrieval failure where a word feels imminent but cannot be immediately produced. This study tracked young and older adults for 28 days using TOT retrospective questionnaires and diary methods to examine the characteristics and mechanisms of cognitive aging in TOT within natural contexts. Results showed: (1) TOT exhibits cognitive aging effects in natural contexts, with older adults experiencing higher TOT frequency than young adults; (2) During TOT episodes, individuals retrieve alternative words and semantically related information, with young adults reporting stronger metacognitive experiences of excitement and fatigue than older adults; (3) Target words were almost always retrieved after TOT occurrence, though older adults required longer resolution times than young adults, with no age difference in resolution rates. TOT resolution time was influenced by cognitive and metacognitive factors including target word familiarity, presence of alternative words, and subjective experiences. (4) Recording TOT cognitive and metacognitive states affected subsequent TOT states and prolonged resolution times, indicating that participants invested greater effort and time in resolving TOT. Both cognitive and metacognitive factors influenced TOT resolution time.

Keywords: TOT; diary study; aging; cognitive perspective; metacognitive perspective

1. Introduction

In daily life, we often experience the phenomenon of having a word “on the tip of our tongue” –unable to immediately articulate something we intend to say. This phenomenon is known as the tip-of-the-tongue (TOT) state (Brown & McNeill, 1966; Burke, Mackay, Worthley, & Wade, 1991). TOT represents a word retrieval failure during speech production that occurs in both children and adults (Hanly & Vandenberg, 2010), particularly among older adults who experience TOT almost daily (Brown, 2012), indicating age-related cognitive decline in speech production. When TOT occurs, speech fluency is disrupted, accompanied by pauses and retrieval errors that impede social communication and indirectly compromise physical and mental well-being (Trull & Phares, 2001). However, research on the mechanisms underlying this phenomenon, especially regarding its causes in natural contexts, remains scarce. Investigating TOT can illuminate the cognitive mechanisms of speech production and provide guidance for preventing, delaying, or even ameliorating age-related speech production decline.

1.1 Research Perspectives and Influencing Factors in TOT Cognitive Aging

Research indicates that during TOT episodes, people can generally retrieve semantic information about target words (Fieder, Nickels, & Biedermann, 2014)

but cannot or can only partially retrieve phonological information (Cleary, Konkel, Nomi, & McCabe, 2010). For example, when experiencing TOT for a familiar person's name, individuals can typically retrieve semantic information such as gender, occupation, and hobbies, but cannot immediately produce the full name. Second, TOT episodes are typically accompanied by conscious metacognitive experiences, including emotions such as tension, anxiety, and excitement, as well as a sense of impending retrieval, which can effectively predict subsequent successful target word retrieval (Schwartz, 2002). After a period of time, people usually successfully retrieve target words, with success rates reaching 95% (Burke et al., 1991). Thus, TOT involves both speech production difficulties and strong metacognitive experiences. Researchers have explained this phenomenon from both cognitive and metacognitive perspectives.

From the cognitive perspective, TOT reflects information retrieval failure in speech production, with aging further increasing this likelihood, typically explained by the transmission deficit hypothesis and the blocking hypothesis. Specifically, the transmission deficit hypothesis (Burke et al., 1991) posits that aging weakens connections between semantic and phonological nodes of target words, preventing effective transmission of activation from semantic to phonological nodes and resulting in phonological retrieval failure. The blocking hypothesis (Jones, 1989) suggests that activation and retrieval of non-target words semantically or phonologically related to the target word temporarily block target word retrieval, with the root cause being older adults' declining ability to inhibit non-target words (Hasher & Zacks, 1988; Stoltzfus, Hasher, & Zacks, 1996). While these theories partially explain TOT cognitive aging—for instance, both account for partial information retrieval, particularly semantic information—they cannot explain the conscious metacognitive experiences accompanying TOT or the high retrieval success rates following TOT. These theories overlook the role of self-monitoring mechanisms (Levelt, Roelofs, & Meyer, 1999) in speech production. Cognitive perspective research has found that TOT-influencing factors primarily relate to target word linguistic properties and individual characteristic variables. For example, proper nouns (e.g., names) are more likely to produce TOT than other nouns (Cohen & Faulkner, 2011), phonological primes related to target words can effectively reduce TOT occurrence and facilitate resolution (Abrams & Rodriguez, 2005; Farrell & Abrams, 2011; Pureza, Soares, & Comesaña, 2013), nouns and low-frequency words are more prone to TOT (Hanly & Vandenberg, 2010), and syllable frequency (Farrell & Abrams, 2011) and phonological neighborhood density (Sadat, Martin, Costa, & Alario, 2014) significantly affect TOT occurrence. Additionally, older adults' declining inhibitory ability leads to more TOT episodes (Peng & Mao, 2018).

Metacognitive perspective research focuses on individuals' experiences and states during TOT episodes, emphasizing the positive role of experience in speech production. According to metacognitive theory of TOT, during speech production the metacognitive system performs heuristic reasoning about target word retrievability based on various perceived cues (Schwartz & Metcalfe, 2011), monitoring

the retrievability of target words in memory systems. TOT occurrence reflects a state where the target word exists in memory but cannot be temporarily retrieved. As a metacognitive monitoring judgment, TOT influences metacognitive control systems, which regulate specific cognitive processes of speech production based on monitoring results, thereby achieving successful target word retrieval after TOT. Indeed, both older and young adults almost always successfully retrieve target words after TOT, with older adults showing slightly higher success rates than young adults (Schwartz, 2002), reflecting that older adults' monitoring and control abilities in speech production are not compromised by aging. TOT-accompanying metacognitive experiences may participate in regulating speech production processes alongside metacognitive monitoring and control functions, with experiences driving and guiding TOT regulation activities, while regulation activities such as monitoring may generate metacognitive experiences (Wang & Guo, 2000). Unlike cognitive explanations, the metacognitive TOT explanation suggests older adults maintain high monitoring and control capabilities in speech production. Metacognitive perspective factors influencing TOT are primarily non-linguistic attributes, such as perceptual cues–cue familiarity (Cleary et al., 2014; Schwartz & Smith, 1997), number of cues (Schwartz & Smith, 1997), and emotions elicited by cues (D' Angelo & Humphreys, 2012; Schwartz, 2010). Additionally, researchers have found metacognitive experiences can effectively predict subsequent successful target word retrieval. For example, TOT episodes accompanied by emotions show significantly higher resolution rates than emotionless TOT (Schwartz, Travis, Castro, & Smith, 2000), with longer resolution times (Schwartz, 2001). High-arousal TOT episodes also show higher subsequent retrieval success rates (Kuipers, 2013; Schwartz, Travis, Castro, & Smith, 2000).

In summary, TOT research on speech production has adopted both cognitive and metacognitive perspectives. Cognitive perspective research emphasizes linguistic properties and general cognitive abilities (e.g., inhibiting irrelevant information), while metacognitive perspective research emphasizes individuals' experiential states during TOT, including emotional states and metacognitive experiences of cues. No study has comprehensively examined how cognitive and metacognitive factors jointly influence TOT resolution. Burke et al. (1991) documented metacognitive experiences accompanying TOT but did not consider their intrinsic relationships with TOT occurrence.

1.2 Comparing Natural and Laboratory Contexts

Most TOT research has been conducted in strictly controlled laboratory settings using standard paradigms (e.g., word definition and picture naming paradigms) (D' Angelo & Humphreys, 2015; Wilhelm & Joolingen, 2013; Souchay & Smith, 2013; Buján, Galdo-Álvarez, Lindín, & Díaz, 2012), which can reveal both age-related characteristics and how target word features affect TOT processes. Nevertheless, researchers have found substantial differences between TOT performance in laboratory and natural contexts. First, target

word retrieval success rates after TOT in natural contexts exceed 95%, with older adults showing slightly higher resolution rates than young adults (Burke et al., 1991; Schwartz, 2002), whereas laboratory studies show only about 40% resolution rates (Schwartz, 1999). If relying solely on laboratory research, we might conclude TOT represents a special type of word retrieval failure where most target words cannot be retrieved. However, considering the high resolution rates in natural contexts, TOT cannot be simply viewed as word retrieval failure in speech production. Second, laboratory research often neglects the positive role of individuals' subjective emotional experiences in resolving TOT (Wang & Guo, 2000). Third, social cues in natural contexts can indirectly affect TOT occurrence by influencing social motivation and expectations (Schwartz, 2002), which laboratory research cannot easily examine.

Compared to experimental methods, diary methods in natural contexts collect daily life data to “capture life as it is lived” (Bolger et al., 2003). These methods are simple and versatile, employing paper-pencil, handheld computers, online questionnaires, and audio recordings. Diary studies are event-based, allowing participants to promptly record event-related questions and experiences, reducing retrospective bias (Reis & Gable, 2000). Diary data are multilevel nested, with each recording nested within individuals, enabling examination of both within-person effects (between recordings) and between-person differences, as well as time-lagged effects (Bolger, Davis, & Rafaeli, 2003) that consider how previous cognitive and behavioral states affect subsequent recordings. Furthermore, diary methods allow examination of variable changes over time, relationships between short-term experiential states and behaviors, and relationships between person or situational characteristics.

The TOT diary method requires individuals to immediately record TOT occurrences during a specified period in natural contexts, including individual characteristics, cognitive information retrieved during TOT (semantic and phonological information, non-target words, retrieval strategies), and accompanying metacognitive experiences (e.g., tension, frustration) (Burke et al., 1991). Burke et al. (1991) used diary methods to investigate TOT aging mechanisms and influencing factors in natural English speech production, finding that older adults produced more TOTs than young adults (especially for proper names like people's names) and fewer non-target words. Regarding TOT resolution, no significant age differences emerged, though older adults more frequently used passive “pop-out” strategies allowing target words to emerge spontaneously (Burke et al., 1991).

1.3 Different Models of Speech Production: Cross-Linguistic Differences

Existing TOT cognitive aging research has focused exclusively on alphabetic languages (e.g., English and Dutch). Alphabetic and non-alphabetic languages (e.g., Chinese) differ in orthographic structure and spelling rules, which may lead to different cognitive characteristics and prevent simple generalization of

alphabetic language findings to non-alphabetic languages. Research shows that in alphabetic language reading, people easily extract whole-word phonology, whereas Chinese, as a logographic script, facilitates direct whole-word semantic extraction (Ho, Law, & Ng, 2000). More importantly, Chinese exhibits different cognitive processing patterns in word production compared to alphabetic languages.

First, research indicates that the phonological encoding unit in Chinese word production is the syllable (Zhang, 2008; O' Seaghdha et al., 2010; Zhang, Chen, Weekes, & Yang, 2009; Zhang & Weekes, 2009; You, Zhang, & Verdonshot, 2012), whereas in alphabetic languages (e.g., English) it is the phoneme or segment (Schiller, 1998, 1999). Second, the connection pattern between semantic and phonological nodes differs cross-linguistically. In Chinese word production, the temporal course of semantic and phonological information activation is sequential—semantic information activates first, followed by phonological information, with no mutual influence between the two (Zhu, Damian, & Zhang, 2015; Zhu, Zhang, & Damian, 2016; Zhang, Zhu, & Damian, 2018). In contrast, alphabetic language speech production shows overlapping temporal courses with bidirectional activation and feedback from phonological to semantic nodes (Starreveld & La Heij, 1995; Dell' Acqua et al., 2010). Compared to alphabetic languages, connections between semantic and phonological nodes may be weaker in Chinese, with different patterns of relationship between semantic and phonological processing across languages. According to the cognitive perspective on TOT, TOT primarily results from deficits or insufficient activation in phonological retrieval, likely leading to cross-linguistic differences in TOT cognitive aging.

In summary, first, although existing experimental research can explain TOT aging mechanisms and influencing factors from both cognitive and metacognitive perspectives, these perspectives yield divergent conclusions: the cognitive perspective suggests TOT aging indicates speech production decline, whereas the metacognitive perspective suggests older adults do not decline in metacognitive aspects of speech production. Most studies have examined TOT from only one perspective. Burke et al. (1991) documented metacognitive experiences accompanying TOT but did not consider their intrinsic relationships with TOT occurrence. Second, most research has used experimental methods with low ecological validity, failing to examine how physical and mental states and social cues in natural contexts affect speech production. Third, existing research has focused on alphabetic languages, yet Chinese speech production patterns differ fundamentally from alphabetic languages.

Therefore, this study employed high-ecological-validity diary methods to conduct a 28-day longitudinal investigation of TOT aging in natural contexts, examining daily TOT occurrences among young and older Chinese adults. Recordings included both cognitive and metacognitive aspects of TOT, with analysis of occurrence mechanisms and influencing factors through ANOVA and regression analysis. Multilevel data analysis examined how TOT recordings affected

subsequent occurrences and how metacognitive states influenced cognitive processes. Based on the transmission deficit hypothesis and blocking hypothesis from the cognitive perspective, and monitoring and control system concepts from the metacognitive perspective, combined with previous findings, we hypothesized that TOT aging results from joint effects of factors from both perspectives. Specifically, from the cognitive perspective, increased TOT may relate to weakened phonological activation in older adults' speech production; from the metacognitive perspective, older adults' monitoring and control abilities may not decline with age, helping them resolve TOT states and discover target words. We verified these hypotheses through multiple analytical methods (ANOVA, regression, multilevel analysis).

2. Method

2.1 Participants

We recruited 31 young adults (16 males, 15 females) through university posters, student websites, and WeChat public accounts, and 36 older adults (13 males, 23 females) through university posters and visits to university communities and parks. All participants had normal or corrected-to-normal vision and normal hearing. After data collection, we assessed older adults' general cognitive ability using the Chinese version of the Montreal Cognitive Assessment Scale (MoCA). Older adults scoring below 26 were excluded from statistical analysis (Sörös, Bose, Sokoloff, Graham, & Stuss, 2011). Following MoCA assessment, three female older adults were excluded, leaving 33 older adults with a mean MoCA score of 27.59 ± 1.27 . Demographic information appears in Table 1. Educational years did not differ significantly between young and older adults ($t(60) = 1.19, p = 0.24$). Participants received compensation upon completion.

Table 1 Age and Educational Years for Young and Older Groups (M \pm SD)

Group	Age Range	Educational Years
Young	18-25	15.50 ± 1.71
Older	60-81	16.33 ± 2.10

Note: Educational years (years)

2.2 Research Tools

Research tools consisted of two parts: a retrospective questionnaire and a TOT diary booklet.

Retrospective Questionnaire. We developed a semi-structured questionnaire based on Burke et al. (1991), comprising nine items covering basic personal information and TOT occurrences during the previous week. The latter section

included TOT frequency, target word part of speech, presence of related alternative words, TOT resolution strategies, resolution probability, and subjective experiences (see Appendix 1).

TOT Diary Booklet. We designed the diary booklet based on Burke et al.'s (1991) diary method (see Appendix 2). The booklet contained 12 items for real-time recording of the entire TOT process from occurrence to resolution over four consecutive weeks. The booklet comprised four basic dimensions: (1) Target word basic characteristics: Item 2 (word type: name, place, object, abstract noun, movie/TV/book title, other), Item 3 (for names: years known, contact frequency per year, time since last contact), and Item 4 (familiarity rating on a 7-point scale); (2) Related information retrieval during TOT: Item 5 (information retrievable during TOT) and Item 7 (alternative words considered); (3) TOT resolution process: Item 1 (occurrence time), Item 11 (resolution time), Item 6 (retrieval strategies), Item 10 (final resolution strategy: repeated thinking, asking others, checking references, doing nothing), Item 8 (whether target word was retrieved), and Item 9 (correct target word); (4) Subjective emotional experiences during TOT: tension, excitement, fatigue, and comfort levels (Item 12). Dimensions 1-3 represented cognitive perspective indicators, while Dimension 4 represented metacognitive perspective indicators.

2.3 Procedure

This study employed a combination of semi-structured retrospective questionnaires and TOT diary booklets to assess TOT occurrences across age groups. The procedure was as follows: Step 1, before formal data collection, the experimenter described the TOT state to ensure participants understood its basic characteristics, introduced diary booklet requirements, provided training, and had participants complete the retrospective questionnaire. Step 2, participants recorded TOT occurrences in real-time over one month. During this period, the experimenter reminded participants weekly via telephone interviews, inquiring about completion status and answering questions to ensure understanding of requirements. Step 3, the experimenter collected and reviewed each participant's diary booklet to ensure clear, legible recordings, and finally administered the MoCA test to older adults.

3. Results

One young adult dropped out, leaving 30 participants (16 males, 14 females). One older adult was excluded due to incomplete diary records, leaving 32 older adults (10 males, 22 females)¹.

¹ We used G*Power 3 to calculate minimum sample size. For independent-samples t-tests ($\alpha = 0.05$, $\beta = 0.2$, large effect size = 0.8), required sample size was 54. For two-way repeated-measures ANOVA ($\alpha = 0.05$, $\beta = 0.2$, medium effect size = 0.25), required sample size was 34. Our sample size of 62 met these requirements.

Statistical analyses of retrospective questionnaire and diary booklet results follow.

3.1 Retrospective Questionnaire Coding and Results

“Frequency of occurrence,” “estimated probability of recall,” and “degree of worry” were continuous data. “Estimated target word type,” “retrievable information” (coded as semantic, phonological, or lexical information), and “target word retrieval methods” were categorical data scored as 0 or 1 (1 = belongs to category, 0 = does not belong).

Retrospective questionnaire results showed no age differences in recalled TOT frequency over the past month ($t(61) = -1.05$, $p = 0.30$), TOT resolution rate (successful target word retrieval) ($t(61) = -0.51$, $p = 0.61$), or subjective experience (degree of worry) ($t(61) = -0.38$, $p = 0.71$).

3.2 TOT Diary Booklet Coding and Results

We conducted five analyses on this data. First, we analyzed basic target word characteristics during TOT, including familiarity, occurrence frequency, and word type. Second, we examined related information retrieval during TOT, including recall characteristics (semantic/phonological), alternative words, and physical/mental states. Third, we analyzed resolution strategies and final resolution approaches. Fourth, we used regression analysis to examine factors influencing TOT resolution time. Fifth, we used multilevel data analysis to investigate how recording TOT states affected subsequent TOT resolution time. Our analyses focused on whether young and older adults showed different patterns—namely, characteristics and influencing factors of TOT cognitive aging.

3.2.1 Age Effects on Target Word Characteristics During TOT Table 2 presents target word characteristics during TOT, including familiarity judgments, occurrence counts, and word type features across ages.

Table 2 Target Word Characteristics During TOT by Age ($M \pm SD$)

Measure	Young	Older
Total Occurrences	5.50 ± 0.71	6.41 ± 2.96
Total Resolved	4.83 ± 2.10	5.26 ± 1.15
Familiarity	4.83 ± 0.71	4.83 ± 0.71
Abstract Words	0.47 ± 0.27	0.10 ± 0.13
Movie/TV/Book Titles	0.14 ± 0.21	0.16 ± 0.20
Proper Nouns	0.11 ± 0.22	0.02 ± 0.05
Non-proper Nouns	0.10 ± 0.13	0.14 ± 0.19
Proper Noun Proportion	0.78 ± 0.23	0.82 ± 0.21
Non-proper Noun Proportion	0.21 ± 0.34	0.17 ± 0.25

Note: Familiarity: 1 = very unfamiliar, 7 = very familiar; Word types are proportions.

Target Word Familiarity. Across 62 participants, 350 TOT episodes occurred, with 326 (93%) successfully resolved. Mean target word familiarity ratings (averaged per participant) appear in Table 2. An independent-samples t-test revealed no age difference in target word familiarity ($t(60) = 1.01, p = 0.31$).

TOT Occurrence Frequency. We calculated total TOT occurrences and resolutions per participant. An independent-samples t-test showed a significant age difference in occurrence frequency, $t(60) = -2.40, p = 0.02$, with young adults ($M = 4.83$) experiencing fewer TOTs than older adults ($M = 6.41$).

Target Word Categories. We calculated occurrence counts and proportions per participant. Analysis revealed that name retrieval produced the highest TOT proportion (53%), followed by movie/TV/book titles (15%), place names (12%), object names (14%), and abstract words (6%). Figure 1 [Figure 1: see original paper] shows the percentage of different word types producing TOT for young and older adults. A 2 (age: young, older) \times 5 (word type: name, place, object, abstract, movie/TV/book) mixed-design ANOVA on TOT proportions revealed violated sphericity ($p < 0.001$), corrected using Greenhouse-Geisser. Significant main effects emerged for word type ($F(4, 240) = 48.89, p < 0.001, p^2 = .45$) and age ($F(1, 60) = 4.55, p < 0.05, p^2 = 0.07$), with a marginally significant interaction ($F(4, 240) = 2.37, p = 0.05, p^2 = 0.04$). Simple effects analysis showed a significant age effect for abstract words, with young adults ($M = 0.11$) showing significantly higher TOT probability for abstract words than older adults ($M = 0.02$) ($F(1, 60) = 11.37, p = 0.001, p^2 = 0.16$).

We further combined names, place names, and movie/TV/book titles into proper nouns, and object names and abstract words into non-proper nouns. A 2 (age) \times 2 (word type: proper vs. non-proper) mixed-design ANOVA revealed violated sphericity ($p < 0.001$), corrected using Greenhouse-Geisser. Significant main effects emerged for word type ($F(1, 120) = 201.73, p < 0.001, p^2 = 0.77$), with proper nouns ($M = 0.79$) producing significantly higher TOT probability than non-proper nouns ($M = 0.10$), and for age ($F(1, 60) = 4.55, p < 0.05, p^2 = 0.07$). The age \times word type interaction was not significant ($F(1, 120) = 1.66, p = 0.20$).

Figure 1 Age differences in TOT occurrence across different word types

Further analysis of name-retrieval TOTs examined how years known, contact frequency, and recency (time since last contact) affected TOT occurrence. Due to high variability, we log-transformed the data. Independent-samples t-tests revealed a significant age difference in “years known” ($t(37) = -5.00, p < 0.001$), with older adults knowing or contacting names for significantly longer ($M = 1.08$) than young adults ($M = 0.34$). Contact frequency and recency showed no age differences ($t_s = -0.33, p_s = 0.45$).

3.2.2 Age Effects on Related Information Retrieval During TOT Table 3 presents young and older adults' assessments of retrieved information characteristics during TOT (means and standard deviations).

Table 3 Characteristics of Retrieved Information During TOT by Age (M \pm SD)

Information Type	Young	Older
Target Word Info	0.96 \pm 0.09	0.97 \pm 0.13
Alternative Word Info	0.06 \pm 0.12	0.07 \pm 0.13
Alternative Word Rate	0.87 \pm 0.24	0.87 \pm 0.13
Semantic Info	0.58 \pm 0.27	0.58 \pm 0.23
Phonological Info	0.20 \pm 0.21	0.27 \pm 0.24
Tension	3.57 \pm 1.14	3.16 \pm 1.19
Fatigue	4.08 \pm 0.88	3.09 \pm 1.20
Excitement	4.10 \pm 0.91	3.09 \pm 1.20
Comfort	3.85 \pm 0.93	4.11 \pm 1.22

Note: Target word and alternative word information are proportions; Physical/mental states rated 1-7 (1 = not at all, 7 = very much).

Semantic and Phonological Information Retrieval. TOT episodes involved retrieval of target word-related information, which we categorized as semantic and phonological (see Table 3). We calculated each participant's proportion of semantic and phonological information retrieval per TOT episode. A 2 (age: young, older) \times 2 (information type: semantic, phonological) mixed-design ANOVA revealed a significant main effect for information type ($F(1, 60) = 1915.07, p < 0.001, \eta^2 = 0.97$), with participants retrieving more semantic information ($M = 0.97$) than phonological information ($M = 0.07$) during TOT. No other main effects or interactions were significant ($F_s \leq 0.13, p_s \geq 0.72$).

Characteristics of TOT Alternative Words. We similarly categorized alternative word information as semantic and phonological and calculated each participant's retrieval proportions (see Table 3). An independent-samples t-test on alternative word quantity revealed no age difference ($t(60) = 0.07, p = 0.94$). Comparing alternatives to target words, we classified them as semantically or phonologically related. A 2 (age) \times 2 (alternative type: semantic, phonological) mixed-design ANOVA revealed a significant main effect for alternative type ($F(1, 60) = 51.87, p < 0.001, \eta^2 = 0.46$), with semantically related alternatives ($M = 0.58$) significantly more frequent than phonologically related alternatives ($M = 0.23$). No other main effects or interactions were significant ($F_s \leq 0.99, p_s \geq 0.32$).

Physical/Mental State Ratings². We calculated each participant's mean ratings for "physical/mental state during TOT" and conducted independent-samples t-tests on subjective metacognitive experiences (tension, excitement, fatigue, comfort). Results showed no age differences in tension ($t(60) = 1.37, p$

= 0.18) or comfort ($t(60) = -0.96, p = 0.35$). However, significant age differences emerged for fatigue ($t(60) = 3.68, p < 0.001$), with young adults ($M = 4.08$) reporting greater fatigue than older adults ($M = 3.09$), and for excitement ($t(60) = 3.72, p < 0.001$), with young adults ($M = 4.10$) showing higher excitement than older adults ($M = 3.09$). See Figure 2 [Figure 2: see original paper].

² Previous studies have not reported the reliability or validity of this questionnaire and diary booklet. We conducted test-retest reliability analysis using our data. Test-retest reliability refers to consistency across repeated administrations of the same scale to the same participants, calculated as the Pearson correlation between two test administrations, examining cross-time consistency. The physical/mental state ratings in the TOT diary booklet were continuous variables recorded after each TOT episode. We calculated test-retest reliability across these ratings, obtaining a reliability coefficient of 0.62 ($p < 0.01$), indicating good cross-time consistency.

3.2.3 Age Effects on TOT Resolution Process Table 4 presents strategies used by young and older adults during TOT episodes and when attempting target word retrieval.

Table 4 Proportions of Different TOT Resolution Strategies for Young and Older Groups ($M \pm SD$)

Strategy	Young (Final)	Young (Attempt)	Older (Final)	Older (Attempt)
Repeated Thinking	0.47 ± 0.31	0.56 ± 0.24	0.49 ± 0.26	0.50 ± 0.25
Asking Others	0.24 ± 0.28	0.22 ± 0.22	0.25 ± 0.19	0.29 ± 0.22
Checking References	0.23 ± 0.22	0.16 ± 0.16	0.10 ± 0.13	0.08 ± 0.10
Doing Nothing	0.06 ± 0.13	0.06 ± 0.12	0.17 ± 0.26	0.13 ± 0.18

Final TOT Resolution Strategies. We calculated each participant’s strategy counts and proportions (see Table 4). A 2 (age: young, older) \times 4 (resolution strategy: repeated thinking, asking others, checking references, doing nothing) mixed-design ANOVA on final resolution rates revealed violated sphericity ($p < 0.05$), corrected using Greenhouse-Geisser. A significant main effect emerged for resolution strategy ($F(3, 177) = 22.35, p < 0.001, p^2 = 0.28$), with “repeated thinking” ($M = 0.48$) significantly more frequent than other strategies, and “asking others” ($M = 0.25$) significantly more frequent than “doing nothing” ($M = 0.11$). The resolution strategy \times age interaction was marginally significant ($F(3, 177) = 2.19, p = 0.09, p^2 = 0.04$). Simple effects analysis revealed that

young adults ($M = 0.23$) used “checking references” more than older adults ($M = 0.10$) ($F(1, 59) = 8.46, p = 0.005, p^2 = 0.13$), while older adults ($M = 0.17$) used “doing nothing” more than young adults ($M = 0.06$) ($F(1, 59) = 4.43, p = 0.04, p^2 = 0.07$). The age main effect was not significant ($F(1, 59) = 0.10, p = 0.75, p^2 = 0.002$).

Attempted TOT Resolution Strategies. We calculated each participant’s strategy counts and proportions (see Table 4). A 2 (age) \times 4 (attempted strategy) mixed-design ANOVA revealed violated sphericity ($p < 0.001$), corrected using Greenhouse-Geisser. A significant main effect emerged for attempted strategy ($F(3, 180) = 49.62, p < 0.001, p^2 = 0.45$), with “repeated thinking” ($M = 0.53$) significantly more frequent than other strategies, and “asking others” ($M = 0.25$) significantly more frequent than “checking references” ($M = 0.12$) and “doing nothing” ($M = 0.10$). Neither the age main effect ($F(1, 60) = 0.57, p = 0.45, p^2 = 0.01$) nor the age \times strategy interaction ($F(3, 180) = 1.84, p = 0.16, p^2 = 0.03$) was significant.

Figure 2 Physical and mental states during TOT for young and older groups

Figure 3 Comparison of TOT resolution strategies between young and older groups

Effects of Age and Alternative Word Occurrence on Final TOT Resolution. Both young and older adults showed high TOT resolution rates (89.4% and 89.63%, respectively). A 2 (age) \times 2 (presence of alternatives) between-subjects ANOVA revealed no significant effects.

3.2.4 Effects of Cognitive and Metacognitive Factors on TOT Resolution Time: Multiple Linear Regression Analysis Table 5 Multiple Linear Regression Results for TOT Resolution Time

Variable	Step 1	Step 2	Step 3
Age	3.86***	2.63**	2.32*
Alternative Word Presence		-0.16	-0.05
Familiarity		-0.08	-0.08
Tension			-2.90**
Fatigue			2.63**
Excitement			-0.69
Comfort			-1.28
Adjusted R ²			

Note: $p < 0.05$, $p < 0.01$, $p < 0.001$

We calculated TOT resolution time from recorded occurrence and resolution times. Based on previous research and our statistical results, we identified influencing factors including: participant characteristics (age), target word-related stimulus features (familiarity, alternative word presence), and subjective states

during TOT (fatigue, tension, excitement, comfort). Stimulus features represent cognitive factors, while subjective states represent metacognitive factors. To further examine factors influencing TOT resolution time, we conducted multiple linear regression with resolution time as the dependent variable and influencing factors as predictors. Steps were: Step 1 entered age; Step 2 entered alternative word presence and familiarity; Step 3 entered four physical/mental state ratings (tension, fatigue, excitement, comfort). All variables entered using forced entry method.

Regression results showed: Step 1 model was significant ($F(1, 318) = 19.61, p < 0.001$), explaining 6% variance, with age significantly predicting resolution time. Step 2 model was significant ($F(3, 318) = 11.84, p < 0.001$), with familiarity and alternative word presence explaining an additional 4% variance, both significantly correlated with resolution time (negative and positive, respectively). After adding subjective states, Step 3 model remained significant ($F(7, 318) = 6.09, p < 0.001$), explaining an additional 2% variance, with comfort significantly positively correlated with resolution time. The final model explained the most variance (Adjusted $R^2 = 0.10$), indicating TOT involves both cognitive and metacognitive factors (see Table 5).

3.2.5 Effects of Cognitive and Metacognitive Factors on Subsequent TOT: Multilevel Data Analysis In this study, participants recorded TOT occurrences over 28 consecutive days, yielding multilevel nested data with recordings nested within individuals. Considering diary studies' time-lagged effects—examining how previous cognitive and behavioral states affect subsequent cognition and behavior (Bolger, Davis, & Rafaeli, 2003; Ohly, 2010; Eldahan et al., 2016)—we hypothesized that participants' evaluations and resolutions of previous TOT states might affect subsequent TOT occurrence and resolution. In multilevel analyses, we used TOT resolution time as the outcome variable and familiarity (cognitive factor) and excitement, fatigue, comfort, and tension (metacognitive states) as predictors to examine how cognitive and metacognitive states affected resolution time, and how post-TOT states affected subsequent TOT cognitive and metacognitive states. Five models were constructed (1: familiarity, 2: excitement, 3: fatigue, 4: comfort, 5: tension).

Figure 4 [Figure 4: see original paper] presents the two-level model, analyzed using Mplus 8.0: Level 1 (within-person effects) and Level 2 (between-person effects). Level 1 included: effects of cognitive/metacognitive states on resolution time (b1), lagged effects of resolution time on subsequent resolution time (b2), lagged effects of cognitive/metacognitive states on subsequent TOT states (b3), effects of resolution time lag on cognitive/metacognitive states (b4), and lagged effects of cognitive/metacognitive states on resolution time (b5). Level 2 included between-person differences in resolution time and cognitive/metacognitive states.

Multilevel results were complex; we present only significant findings here (full

results in Appendix 3). At Level 1: each recording' s resolution time and familiarity significantly affected subsequent resolution time (estimate = 0.30, $p < 0.01$) and familiarity (estimate = 0.34, $p < 0.01$). Each recording' s resolution time affected familiarity in the next TOT (estimate = -0.14, $p < 0.05$). Each recording' s resolution time and excitement significantly affected subsequent resolution time (estimate = 0.32, $p < 0.01$) and excitement (estimate = 0.18, $p < 0.01$). Each recording' s resolution time and fatigue significantly affected subsequent resolution time (estimate = 0.27, $p < 0.001$) and fatigue (estimate = 0.22, $p < 0.001$). Each recording' s resolution time and tension significantly affected subsequent resolution time (estimate = 0.31, $p < 0.001$) and tension (estimate = 0.13, $p < 0.05$). At Level 2: familiarity and resolution time showed significant autoregressive correlation (estimate = 0.62, $p < 0.05$), indicating higher familiarity was associated with greater autoregressive effects. Other metacognitive states showed no significant between-person effects ($ps > 0.05$). Model 1' s lagged effects of resolution time and familiarity explained 44% and 41% of within-person variance, respectively. Model 2' s lagged effects explained 40% and 36% for resolution time and excitement. Model 3 explained 43% and 31% for resolution time and fatigue. Model 4 explained 42% and 32% for resolution time and comfort. Model 5 explained 44% and 36% for resolution time and tension.

Figure 4 Multilevel data analysis model diagram

Note: Cognitive state: familiarity; Metacognitive states: excitement, fatigue, comfort, tension. b_1 = effect of cognitive/metacognitive states on resolution time; b_2 = lagged effect of resolution time on subsequent resolution time; b_3 = lagged effect of cognitive/metacognitive states on subsequent TOT states; b_4 = effect of resolution time lag on cognitive/metacognitive states; b_5 = lagged effect of cognitive/metacognitive states on resolution time; $\mu_{\text{resolution_time}}$ = mean resolution time (between-person); $\mu_{\text{cognitive_metacognitive}}$ = mean familiarity/excitement/fatigue/comfort/tension (between-person)

Multilevel results indicated that post-TOT cognitive and metacognitive states affected subsequent TOT cognitive and metacognitive states, and each TOT' s resolution time affected the next resolution time.

4. Discussion

Using high-ecological-validity diary methods, this study examined TOT mechanisms and influencing factors in natural contexts. Results demonstrated cognitive aging effects in Chinese speech production, with older adults experiencing more TOTs than young adults. Both groups retrieved more semantic than phonological information and produced equivalent numbers of alternative words, with no age differences. Both groups achieved similarly high resolution rates (near 90%), but used different resolution strategies: young adults preferred active strategies, while older adults used spontaneous emergence strategies. Young adults experienced higher fatigue and excitement than older adults during TOT.

Older adults required longer resolution times than young adults, with resolution time influenced by target word familiarity, alternative word presence, and subjective experiences. Multilevel modeling revealed that assessments of post-TOT familiarity and physical/mental states affected subsequent resolution times, indicating both cognitive and metacognitive factors influence speech production.

4.1 Effects of Age and Target Word Type on TOT Occurrence

This study found age-related effects in natural speech production, with older adults experiencing significantly more TOTs than young adults. Analysis of target word familiarity and years known revealed that older adults had known or contacted target names longer than young adults, though familiarity did not differ significantly. This aligns with the transmission deficit hypothesis (Burke et al., 1991), which posits that aging weakens connections between semantic and phonological nodes, preventing sufficient transmission of semantic activation to phonological nodes and impairing phonological retrieval. In other words, age affects phonological but not semantic retrieval (Gollan & Brown, 2006). The absence of age differences in familiarity likely reflects judgments based on semantic information. Weakened semantic-phonological connections in older adults make these words more susceptible to TOT.

Although no age differences emerged in proper noun TOT rates, proper nouns (names) showed significantly higher TOT rates than object names and abstract words, consistent with previous research. Burke et al. (1991) found proper nouns accounted for approximately 74% of natural TOTs. Laboratory studies also show proper noun retrieval produces more TOT (Evrard, 2002). Researchers suggest proper nouns, as referential expressions, contain less semantic information than common nouns (Cohen, 1990; Burke et al., 1991; Semenza, 1995; Evrard, 2002), weakening semantic node activation and reducing connections to associated words in memory, making proper nouns more difficult to retrieve and producing more TOT. Young adults produced more abstract word TOTs than older adults, possibly because young university students encounter books and novelties more frequently, using abstract words more often, reflecting pragmatic differences in language use (Burke et al., 1991).

Additionally, natural TOT word types included not only the most frequent nouns (53%) but also place names (12%), object names (13%), movie/TV/book titles (15%), and abstract words (7%). Previous experimental studies mostly induced noun TOTs, whereas our naturalistic study yielded richer spontaneous TOT categories, providing authentic materials for future experimental research.

4.2 Characteristics of Information Retrieval During TOT

During TOT, people typically retrieve partial target word information and produce alternative words. This study found no age differences in alternative word quantity between Chinese older and young adults, unlike findings in English. Burke et al. (1991) and Cohen and Faulkner (2011) found young adults

(and middle-aged adults) produced more alternatives than older adults during TOT. These findings do not support blocking or transmission deficit hypotheses (Hasher & Zacks, 1988; Jones, 1989; Stoltzfus, Hasher, & Zacks, 1996), which suggest alternatives block target retrieval and that older adults' larger vocabularies (Ben-David, Erel, Goy, & Schneider, 2015) activate more non-target words, producing more TOT. Since we ensured normal general cognitive ability in older adults via MoCA, we speculate that naturalistic TOT cognitive aging does not result from declined inhibition of alternatives.

Second, most alternatives were semantically related to target words. For example, when experiencing TOT for a name, individuals could retrieve semantic information like occupation, gender, and appearance. This indicates semantic information retrieval during TOT. We found phonologically related information ($M = 0.07$) was retrieved far less than semantically related information ($M = 0.96$), consistent with the transmission deficit hypothesis (Burke et al., 1991). Research showing that phonologically related primes reduce TOT occurrence and facilitate resolution (Abrams & Rodriguez, 2005; Farrell & Abrams, 2011) confirms that weakened phonological activation causes TOT.

Third, no age differences emerged in phonological alternative proportions, unlike alphabetic language findings. In alphabetic languages, Burke et al. (1991) found young adults retrieved more phonological associates (2.07) than older adults (1.53), and Cohen and Faulkner (2011) found higher phonological retrieval rates in young adults (0.28) than older adults (0.22). Different TOT cognitive aging patterns between Chinese and alphabetic languages may relate to different speech production models. Alphabetic languages show interactive activation between semantic and phonological stages with bidirectional transmission and multiple phonological activations of non-target items. With aging, older alphabetic language speakers show stronger semantic-phonological connections and activate more semantically related non-target words (Dell'Acqua et al., 2010; Starreveld & La Heij, 1995), producing more alternatives. In contrast, Chinese speech production shows no interaction between semantic and phonological stages, with phonological activation only for target words (Zhu et al., 2015, 2016; Zhang et al., 2018). Even if older adults' semantic activation does not decline, it does not spread to semantically related non-target words, resulting in equivalent alternative production between Chinese older and young adults.

Finally, metacognitive experience ratings during TOT showed young adults experienced more fatigue and excitement than older adults. According to the metacognitive perspective, TOT-accompanying experiences reflect metacognitive system monitoring and control of speech production (Schwartz & Metcalfe, 2011). Specifically, after TOT occurs, individuals adopt different resolution strategies to control speech production based on metacognitive experiences, aiming for successful target retrieval. Differences in emotional experiences may relate to different retrieval strategies (Kuipers, 2013; Schwartz, 2008; Warriner & Humphreys, 2008). Compared to older adults, young adults may adopt more

active strategies—we found they used “checking references” more frequently, while older adults used “doing nothing” more often, allowing spontaneous emergence. Consequently, young adults’ TOT resolution times were significantly shorter than older adults’, consistent with previous research (Burke et al., 1991). Meanwhile, active resolution strategies may consume more attentional resources, leading young adults to experience greater fatigue. Higher excitement in young adults reflects their ability to predict imminent target retrieval after active strategy use, generating more positive emotions. Schwartz (2001) found similar patterns: TOT episodes accompanied by excitement enabled retrieval of more semantic information (e.g., part of speech, number) compared to those accompanied by frustration.

Our results showed older adults experienced more TOTs but lower ratings on some metacognitive experiences (fatigue, excitement) than young adults, while other experiences did not differ, suggesting TOT frequency is unrelated to metacognitive experiences. TOT does not occur continuously due to metacognitive experiences; rather, specific metacognitive experiences accompany TOT episodes. Different patterns across experience dimensions suggest differential sensitivity between groups. During questionnaire collection interviews, we found young adults paid more attention to physical/mental states, while older adults may have shown recording bias due to forgetting. Previous research indicates older adults can monitor and control target word retrieval with success rates comparable to young adults (Burke et al., 1991). In our study, although older adults’ physical/mental state ratings were lower than young adults’, their TOT resolution rate reached 94%, slightly higher than young adults’ 92%, suggesting older adults’ speech production monitoring and control may be equivalent to young adults’, possibly assisted by metacognitive state evaluations. This highlights the need to examine relationships between TOT cognitive and metacognitive factors and their interactive effects on resolution time and strategies. Our regression and multilevel analyses attempted this, but rigorous laboratory studies are needed for deeper investigation.

4.3 Factors Influencing TOT Resolution

First, resolution strategies affect target word retrieval after TOT. Both groups achieved high resolution rates (89.4% and 89.63%). However, groups differed in strategies: young adults used “checking references” more, while older adults used “doing nothing” spontaneous strategies. These differences may relate to knowledge or vocabulary size. Older adults have larger vocabularies than young adults (Ben-David, Erel, Goy, & Schneider, 2015), and larger vocabularies produce fluctuations in word accessibility with dual effects: they increase TOT likelihood but also enable spontaneous retrieval when target words reach high activation intensity. During word production, when target word activation is low, retrieval fails or TOT occurs; when activation is high, successful retrieval occurs. Older adults can spontaneously retrieve when activation is high, thus using “doing nothing” more frequently.

Second, target word familiarity affected resolution time, consistent with transmission deficit predictions: familiarity and usage frequency influence connection strength between semantic and phonological representations. Greater familiarity and frequency strengthen node connections, enabling rapid semantic-to-phonological transmission and shortening TOT resolution time (Burke et al., 1991). Additionally, alternative word presence significantly predicted resolution time, with longer resolution when alternatives were present. This suggests alternatives may compete with target words, interfering with phonological retrieval and prolonging resolution, supporting the blocking hypothesis (Abrams & Rodriguez, 2005; Jones, 1989; White et al., 2013). Although we found no effect of alternatives on TOT occurrence, they affected resolution time, particularly semantically related alternatives.

Notably, subjective experiences also influenced resolution time. Specifically, higher comfort during TOT predicted longer resolution time. While previous research did not directly examine this relationship, it found subjective experiences predict resolution. For example, TOT urgency effectively predicted resolution, with urgent TOTs showing five times higher resolution rates than non-urgent TOTs (Schwartz, Travis, Castro, & Smith, 2000). Similarly, Schwartz (1999) found TOTs with emotional experiences had lower resolution rates than emotionless TOTs in natural contexts. Different subjective experiences may reflect different metacognitive regulation of retrieval. Individuals with high comfort may be less motivated to retrieve targets, showing weaker resolution motivation and investment than low-comfort individuals (Schwartz, 2001, 2008), resulting in longer resolution times.

Multilevel analyses revealed that within-person recordings of TOT cognitive and metacognitive states affected subsequent resolution time and experiences, and each TOT's resolution time increased the next resolution time. Using resolution time as the cognitive variable, we found each metacognitive experience (excitement, fatigue, comfort, tension) increased subsequent resolution time, reflecting metacognitive effects on cognitive processes. This suggests that recording and evaluating cognitive and metacognitive states after TOT increases resolution probability and enhances confidence, leading participants to invest more effort and time in resolving subsequent TOTs. This speculation requires laboratory verification.

This diary study investigated naturalistic TOT patterns and mechanisms across age groups with high ecological validity and authentic responses, overcoming laboratory study limitations. However, limitations remain. First, naturalistic diary studies cannot control extraneous variables, and subjective reports may contain bias (Burke, MacKay, Worthley, & Wade, 1991; Cohen & Faulkner, 2011). We used paper-pencil diaries, which, while convenient for immediate recording, cannot eliminate forgetting risk or ensure strict compliance with recording every TOT state—a common issue in diary studies (Bolger, Davis, & Rafaeli, 2003). Future research should combine rigorous laboratory studies controlling specific variables to explore how semantically and phonologically related non-

target words affect word retrieval aging in speech production.

Comparison with previous results and test-retest reliability analyses indicate our questionnaire and diary booklet possess good reliability and validity. Future laboratory TOT studies can use these measurements to examine how cognitive and metacognitive states affect resolution and strategy use under strict experimental control. For example, in laboratory settings, semantic information may more easily induce Chinese TOT, and providing semantically related alternatives after TOT may shorten resolution time. Naturalistic TOT research provides authentic materials for laboratory studies, allowing researchers to select appropriate TOT stimuli based on specific research purposes.

5. Conclusion

In summary, we found cognitive aging effects in Chinese speech production, with older adults experiencing more TOTs than young adults. Both groups retrieved more semantic than phonological information after TOT and produced equivalent numbers of alternatives, supporting the transmission deficit hypothesis. Young adults experienced stronger metacognitive experiences of excitement and fatigue than older adults after TOT. No age differences emerged in resolution rates, but older adults required longer resolution times than young adults, with resolution time affected by target word familiarity, alternative word presence, and subjective experiences. Assessing TOT states affected subsequent resolution times, indicating that recording and evaluating TOT states led participants to invest more effort and time in resolution. TOT resolution involves both target retrieval cognitive factors and subjective metacognitive experiences, representing joint products of cognitive and metacognitive processes. Future research should use rigorously controlled experimental designs to examine mechanisms of these factors and their interactions.

References

- Abrams, L., & Rodriguez, E. (2005). Syntactic class influences phonological priming of tip-of-the-tongue resolution. *Psychonomic Bulletin & Review*, *12*(6), 1018-1023.
- Ben-David, B. M., Erel, H., Goy, H., & Schneider, B. A. (2015). “Older is always better” : Age-related differences in vocabulary scores across 16 years. *Psychology and Aging*, *30*(4), 856-862.
- Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology*, *54*(1), 579-616.
- Brown, A. S. (2012). *The tip of the tongue state*. Hove, United Kingdom: Psychology Press.
- Brown, R., & McNeill, D. (1966). The “tip of the tongue” phenomenon. *Journal of Verbal Learning and Behavior*, *5*(4), 325-337.

- Buján, A., Galdo-Álvarez, S., Lindín, M., & Díaz, F. (2012). An event-related potentials study of face naming: Evidence of phonological retrieval deficit in the tip-of-the-tongue state. *Psychophysiology*, *49*(7), 980-990.
- 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 & Language*, *30*(5), 542-579.
- Cleary, A. M., Konkell, K. E., Nomi, J. S., & McCabe, D. P. (2010). Odor recognition without identification. *Memory & Cognition*, *38*(4), 452-460.
- Cleary, A. M., Staley, S. R., & Klein, K. R. (2014). The effect of tip-of-the-tongue states on other cognitive judgments. In B. L. Schwartz & A. S. Brown (Eds.), *Tip-of-the-tongue states and related phenomena* (pp. 75-94). New York, NY: Cambridge University Press.
- Cohen, G. (1990). Why is it difficult to put names to faces? *British Journal of Psychology*, *81*(3), 287-297.
- Cohen, G., & Faulkner, D. (2011). Memory for proper names: Age differences in retrieval. *British Journal of Developmental Psychology*, *4*(2), 187-197.
- D' Angelo, M. C., & Humphreys, K. R. (2012). Emotional cues do not increase the likelihood of tip-of-the-tongue states. *Memory and Cognition*, *40*(8), 1331-1338.
- D' Angelo, M. C., & Humphreys, K. R. (2015). Tip-of-the-tongue states reoccur because of implicit learning, but resolving them helps. *Cognition*, *142*, 166-190.
- Dell' Acqua, R., Sessa, P., Peressotti, F., Mulatti, C., Navarrete, E., & Grainger, J. (2010). ERP evidence for ultra-fast semantic processing in the picture-word interference paradigm. *Frontiers in Psychology*, *1*, 177.
- Eldahan, A. I., Pachankis, J. E., Rendina, H. J., Ventuneac, A., Grov, C., & Parsons, J. T. (2016). Daily minority stress and affect among gay and bisexual men: A 30-day diary study. *Journal of Affective Disorders*, *190*, 828-835.
- Evrard, M. (2002). Ageing and lexical access to common and proper names in picture naming. *Brain and Language*, *81*(1), 174-179.
- Farrell, M. T., & Abrams, L. (2011). Tip-of-the-tongue states reveal age differences in the syllable frequency effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(1), 277-285.
- Fieder, N., Nickels, L., & Biedermann, B. (2014). Representation and processing of mass and count nouns: A review. *Frontiers in Psychology*, *5*, 589.
- 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*(3), 462-483.
- Hanly, S., & Vandenberg, B. (2010). Tip-of-the-tongue and word retrieval deficits in dyslexia. *Journal of Learning Disabilities*, *43*(1), 15-19.

- Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In *Psychology of Learning and Motivation* (Vol. 22, pp. 193-225). Academic Press.
- Ho, C. S. H., Law, T. P. S., & Ng, P. M. (2000). The phonological deficit hypothesis in Chinese developmental dyslexia. *Reading and Writing, 13*(1-2), 57-79.
- Jones, G. V. (1989). Back to Woodworth: Role of interlopers in the tip of the tongue phenomenon. *Memory & Cognition, 17*, 69-76.
- Kuipers, S. C. (2013). *Effect of incubation on the resolution of Tip-of-the-Tongue states and the relation with attention and concentration*. Bachelor' s thesis, University of Twente.
- Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*(1), 1-38.
- Ohly, S., Zapf, D., & Dieter. (2010). Diary studies in organizational research: An introduction and some practical recommendations. *Journal of Personnel Psychology, 9*, 79-93.
- O'Seaghdha, P. G., Chen, J.-Y., & Chen, T. M. (2010). Proximate units in word production: Phonological encoding begins with syllables in Mandarin Chinese but with segments in English. *Cognition, 115*(2), 282-302.
- Peng, H. M., & Mao, X. F. (2018). Will the deficit in inhibition increase the rates of tip-of-the-tongue among the elderly? *Acta Psychologica Sinica, 50*(10), 1142-1150. [彭华茂, 毛晓飞. (2018). 抑制对老年人舌尖现象的影响. *心理学报, 50*(10), 1142-1150.]
- Pureza, R., Soares, A. P., & Comesaña, M. (2013). Syllabic pseudohomophone priming in tip-of-the-tongue states resolution: The role of syllabic position and number of syllables. *The Quarterly Journal of Experimental Psychology, 66*, 910-926.
- Reis, H. T., & Gable, S. L. (2000). Event-sampling and other methods for studying everyday experience. In *Handbook of Research Methods in Social and Personality Psychology*.
- Sadat, J., Martin, C. D., Costa, A., & Alario, F. X. (2014). Reconciling phonological neighborhood effects in speech production through single trial analysis. *Cognitive Psychology, 68*, 33-58.
- Schiller, N. O. (1998). The effect of visually masked syllable primes on the naming latencies of words and pictures. *Journal of Memory & Language, 39*(3), 484-507.
- Schiller, N. O. (1999). No role for syllables in English speech production. *Journal of the Acoustical Society of America, 105*(2), 1204.

- Schwartz, B. L. (1999). The phenomenology of naturally-occurring tip-of-the-tongue states: A diary study. In S. Shohov (Ed.), *Advances in Psychology Research* (pp. 71-84). Nova Science Publishers: New York.
- Schwartz, B. L. (2001). The relation of tip-of-the-tongue states and retrieval time. *Memory & Cognition*, *29*(1), 117-126.
- Schwartz, B. L. (2002). *Tip-of-the-tongue states: Phenomenology, mechanism, and lexical retrieval*. *Experimental Psychology*, *49*(3), 147-148.
- Schwartz, B. L. (2008). Working memory load differentially affects tip-of-the-tongue states and feeling-of-knowing judgment. *Memory & Cognition*, *36*(1), 9-19.
- Schwartz, B. L. (2010). The effects of emotion on tip-of-the-tongue states. *Psychonomic Bulletin & Review*, *17*(1), 82-87.
- Schwartz, B. L., & Metcalfe, J. (2011). Tip-of-the-tongue (TOT) states: Retrieval, behavior, and experience. *Memory & Cognition*, *39*(5), 737-749.
- Schwartz, B. L., & Smith, S. M. (1997). The retrieval of related information influences tip-of-the-tongue states. *Journal of Memory and Language*, *36*(1), 68-86.
- Schwartz, B. L., Travis, D. M., Castro, A. M., & Smith, S. M. (2000). The phenomenology of real and illusory tip-of-the-tongue states. *Memory & Cognition*, *28*(1), 18-27.
- Semenza, C. (1995). How names are special: Neuropsychological evidence for dissociable impairment and sparing of proper name knowledge in production. In *Broken memories: Case studies in memory impairment*. Blackwell, Oxford.
- Sörös, P., Bose, A., Sokoloff, L. G., Graham, S. J., & Stuss, D. T. (2011). Age-related changes in the functional neuroanatomy of overt speech production. *Neurobiology of Aging*, *32*(8), 1505-1515.
- Souchay, C., & Smith, S. J. (2013). Subjective states associated with retrieval failures in Parkinson's disease. *Consciousness and Cognition*, *22*(3), 795-805.
- Starreveld, P. A., & La Heij, W. (1995). Semantic interference, orthographic facilitation, and their interaction in naming tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(3), 686-698.
- Stoltzfus, E. R., Hasher, L., & Zacks, R. T. (1996). Working memory and aging: Current status of the inhibitory view. In *Working Memory and Human Cognition* (pp. 66-88).
- Trull, T. J., & Phares, E. J. (2001). *Clinical psychology: Concepts, methods, and profession* (6th ed.). CA: Wadsworth.
- Wang, L., & Guo, D. J. (2000). The nature and components of metacognition. *Acta Psychologica Sinica*, *32*(4), 458-463. [汪玲, & 郭德俊. (2000). 元认知的本质与要素. *心理学报*, *32*(4), 458-463.]

- Warriner, A. B., & Humphreys, K. R. (2008). Learning to fail: Reoccurring tip-of-the-tongue states. *The Quarterly Journal of Experimental Psychology*, *61*(4), 535–542.
- White, K. K., Abrams, L., & Frame, E. A. (2013). Semantic category moderates phonological priming of proper name retrieval during tip-of-the-tongue states. *Language and Cognitive Processes*, *28*, 561–576.
- Wilhelm, P., & Joolingen, W. R. (2013). Effect of incubation on the resolution of tip-of-the-tongue states and the relation with attention and concentration. *Kagaku Kogaku Ronbunshu*, *31*(31), 399–403.
- You, W., Zhang, Q., & Verdonschot, R. G. (2012). Masked syllable priming effects in word and picture naming in Chinese. *Plos One*, *7*(10), e46595.
- Zhang, Q. (2008). Phonological encoding in monosyllabic and bisyllabic Mandarin word production: Implicit priming paradigm study. *Acta Psychologica Sinica*, *40*(3), 253–262. [张清芳. (2008). 汉语单音节和双音节词汇产生中的音韵编码过程: 内隐启动范式研究. *心理学报*, *40*(3), 253–262.]
- Zhang, Q., & Weekes, B. S. (2009). Orthographic facilitation effects on spoken word production: Evidence from Chinese. *Language and Cognitive Processes*, *24*(7/8), 1082–1096.
- Zhang, Q., Weekes, B., Chen, H. C., & Yang, Y. (2009). Independent effects of orthographic and phonological facilitation on spoken word production in Mandarin. *Language and Speech*, *52*, 113–126.
- Zhang, Q., Zhu, X., & Damian, M. F. (2018). Phonological activation of category coordinates in spoken word production: Evidence for cascaded processing in English but not in Mandarin. *Applied Psycholinguistics*, *1–26*.
- Zhu, X., Damian, M. F., & Zhang, Q. (2015). Seriality of semantic and phonological processes during overt speech in Mandarin as revealed by event-related brain potentials. *Brain and Language*, *144*, 16–25.
- Zhu, X., Zhang, Q., & Damian, M. F. (2016). Additivity of semantic and phonological effects: Evidence from speech production in Mandarin. *The Quarterly Journal of Experimental Psychology*, *69*(11), 2285–2304.

Appendix 1: Tip-of-the-Tongue Retrospective Questionnaire

Tip-of-the-tongue: A phenomenon where a word feels imminent but cannot be immediately produced. Characteristics: 1. The content is very familiar; unfamiliar content that cannot be produced does not qualify as TOT. 2. Though not immediately producible, the word's meaning, part of speech, length, and partial phonological information are known. 3. Though not immediately producible, the word can likely be produced later. 4. Though not immediately

producibile, related words are produced during attempts, with awareness that these are not the target word.

I. Basic Personal Information 1. Age: _____ years 2. Gender: A. Male B. Female 3. Education level: A. High school or below B. College or above C. Master' s or above

II. Please respond based on your actual TOT experiences in the past week 1. Rate your TOT frequency over the past month on a 1-7 scale (check one): 1 = never, 7 = very frequent. 2. What types of target words do your TOTs involve? (Check all that apply) A. Names B. Place names C. Object names D. Abstract nouns E. Movie/TV/book titles F. Other 3. What information can you typically recall during TOT? 4. What methods do you generally use to recall target words after TOT? A. Repeatedly thinking about the target word B. Asking others C. Checking references D. Doing nothing 5. What is the probability of eventually recalling target words in daily life? (Circle one) 6. Are you worried about experiencing TOT? Rate on a 1-7 scale (check one): 1 = not worried at all, 7 = very worried.

Appendix 2: Tip-of-the-Tongue Diary Booklet

Please record detailed information for each TOT episode using the form below

Tip-of-the-tongue: A phenomenon where a word feels imminent but cannot be immediately produced. Characteristics: 1. The content is very familiar; unfamiliar content that cannot be produced does not qualify as TOT. 2. Though not immediately producibile, the word' s meaning, part of speech, length, and partial phonological information are known. 3. Though not immediately producibile, the word can likely be produced later. 4. Though not immediately producibile, related words are produced during attempts, with awareness that these are not the target word.

Date: ____ Month ____ Day **Time:** ____ Hour ____ Minute

1. TOT occurrence time
2. Target word type: A. Name B. Place name C. Object name D. Abstract noun E. Movie/TV/book title F. Other
3. If target word is a name, please answer:
 - (1) Years known: _____ years
 - (2) Contact frequency: _____ times/year
 - (3) Time since last contact: _____ years _____ months
4. Target word familiarity (check one): 1 = very unfamiliar, 2 = somewhat unfamiliar, 3 = slightly unfamiliar, 4 = neutral, 5 = slightly familiar, 6 = somewhat familiar, 7 = very familiar
5. What information can you recall during TOT?
6. How did you attempt to recall the target word? A. Repeated thinking B. Asking others C. Checking references D. Doing nothing

7. List all alternative words that came to mind during retrieval attempts
8. Did you finally recall the target word? A. Yes (if yes, answer questions 9-11) B. No (if no, answer questions 12-13)
9. Correct target word: _____
10. Final resolution method: A. Repeated thinking B. Asking others C. Checking references D. Doing nothing
11. Target word recall time (24-hour format): ____ Month ____ Day ____ Hour ____ Minute

12. Physical and mental state during TOT (rate 1-7): - Tension/Frustration: 1 = not at all tense/frustrated, 7 = very tense/frustrated - Fatigue level: 1 = not at all fatigued, 7 = very fatigued - Excitement level: 1 = not at all excited, 7 = very excited - Comfort level: 1 = not at all comfortable, 7 = very comfortable

Appendix 3: Multilevel Data Analysis Results

Model 1: Familiarity and Resolution Time

Parameter	Estimate
Within Level	
Familiarity effect on resolution time	-0.06
Resolution time effect on subsequent resolution time	0.30*
Familiarity effect on subsequent familiarity	0.34***
Familiarity effect on subsequent resolution time	-0.14*
Resolution time effect on subsequent familiarity	-0.004
Between Level	
Familiarity autoregressive correlation	0.62*

Note: $p < 0.05$, $p < 0.01$, $p < 0.001$

Model 2: Excitement and Resolution Time

Parameter	Estimate
Within Level	
Excitement effect on resolution time	0.32***
Resolution time effect on subsequent resolution time	0.18**
Excitement effect on subsequent excitement	-0.07
Excitement effect on subsequent resolution time	-0.10
Resolution time effect on subsequent excitement	-0.45

Model 3: Fatigue and Resolution Time

Parameter	Estimate
Within Level	
Fatigue effect on resolution time	0.27***
Resolution time effect on subsequent resolution time	0.22***
Fatigue effect on subsequent fatigue	-0.02
Fatigue effect on subsequent resolution time	-0.21
Resolution time effect on subsequent fatigue	-0.43

Model 4: Comfort and Resolution Time

Parameter	Estimate
Within Level	
Comfort effect on resolution time	0.31***
Resolution time effect on subsequent resolution time	0.13*
Comfort effect on subsequent comfort	-0.06
Comfort effect on subsequent resolution time	-0.09
Resolution time effect on subsequent comfort	-0.44

Model 5: Tension and Resolution Time

Parameter	Estimate
Within Level	
Tension effect on resolution time	0.30***
Resolution time effect on subsequent resolution time	0.26***
Tension effect on subsequent tension	-0.05
Tension effect on subsequent resolution time	-0.30
Resolution time effect on subsequent tension	-0.05

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv – Machine translation. Verify with original.