

Unraveling the linguistic nature of specific autobiographical memories using a computerized classification algorithm

Keisuke Takano^{1,5} · Mayumi Ueno² · Jun Moriya³ · Masaki Mori⁴ · Yuki Nishiguchi⁴ · Filip Raes¹

Published online: 23 June 2016
© Psychonomic Society, Inc. 2016

Abstract In the present study, we explored the linguistic nature of specific memories generated with the Autobiographical Memory Test (AMT) by developing a computerized classifier that distinguishes between specific and nonspecific memories. The AMT is regarded as one of the most important assessment tools to study memory dysfunctions (e.g., difficulty recalling the specific details of memories) in psychopathology. In Study 1, we utilized the Japanese corpus data of 12,400 cue-recalled memories tagged with observer-rated specificity. We extracted linguistic features of particular relevance to memory specificity, such as past tense, negation, and adverbial words and phrases pertaining to time and location. On the basis of these features, a support vector machine (SVM) was trained to classify the memories into specific and nonspecific categories, which achieved an area under the curve (AUC) of .92 in a performance test. In Study 2, the trained SVM was tested in terms of its robustness in classifying novel memories ($n = 8,478$) that were retrieved in response to cue words that were different from those used in Study 1. The SVM showed an AUC of .89 in classifying the new memories. In Study 3, we extended the binary SVM to a five-class classification of the AMT, which achieved 64%–65% classification accuracy, against the chance level (20%) in the performance tests. Our

data suggest that memory specificity can be identified with a relatively small number of words, capturing the universal linguistic features of memory specificity across memories in diverse contents.

Keywords Autobiographical memory · Natural language processing · Machine learning · Support vector machine

The nature of specific autobiographical memories

The specificity of autobiographical memories has been studied frequently in the past three decades, in the fields of cognitive, social, and clinical psychology. An autobiographical memory (AM) refers to “a memory that is concerned with the recollection of personally experienced past events” (Williams et al., 2007, p. 122). The level of episodic specificity with which such personal memories are recalled has become a target of increased interest related to topics such as aging, cultural differences, and—most importantly—psychopathology (e.g., Addis, Wong, & Schacter, 2008; van Vreeswijk & de Wilde, 2004; Wang, Hou, Tang, & Wiprovnick, 2011). Specificity is important not only for vividly reexperiencing past events, but also for clearly imagining future events. In fact, remembering past events and envisioning possible future events share similar cognitive functions and neural substrates (Addis, Wong, & Schacter, 2007; D’Argembeau & Van der Linden, 2006; Okuda et al. 2003). Thus, AM recall, and particularly the retrieval of episodic specificity, is assumed to play a central role in human functioning because AM specificity contributes to a sense of self and serves as a source for future planning and goal pursuit (Williams et al., 2007).

Reduced AM specificity is considered particularly important in clinical psychology. Empirical studies in this area

✉ Keisuke Takano
Keisuke.Takano@ppw.kuleuven.be

¹ Center for Learning and Experimental Psychopathology, University of Leuven, Tiensestraat 102, Leuven 3000, Belgium

² National Center of Neurology and Psychiatry, Tokyo, Japan

³ Kansai University, Osaka, Japan

⁴ University of Tokyo, Tokyo, Japan

⁵ Nihon University, Tokyo, Japan

typically use the Autobiographical Memory Test (AMT; Williams & Broadbent, 1986), which has been used in more than 100 clinical and nonclinical studies since 1986 (for reviews, see Griffith, Sumner, et al. 2012; Sumner, Griffith, & Mineka, 2010; van Vreeswijk, & de Wilde, 2004; Williams et al. 2007). Such studies have repeatedly shown that people with major depressive disorder (American Psychiatric Association, 2013) find it difficult to retrieve memories of specific personal events (Dalglish & Werner-Seidler, 2014; Williams et al., 2007). In the AMT, participants are asked to recall specific personal memories in response to emotional cue words (e.g., *happy*, *sad*). A *specific* memory refers to an event that occurred on a particular day and did not last longer than a day (e.g., “The day I turned 18, I got new shoes from my girlfriend”). In the AMT, healthy controls are able to provide specific memories on more than 80% of the occasions. However, people with depression tend to respond relatively more often with a memory that summarizes a category of similar events (e.g., “Getting presents from friends and family”; Williams et al., 2007). This lack of specificity, or *overgeneral memory*, is considered a hallmark of depressive cognition. Overgeneral memory is relatively stable over time, despite clinical improvement in the depressive symptoms, since it is associated with poor prognosis in terms of depression and PTSD (Brittlebank, Scott, Williams, & Ferrier, 1993; Kleim & Ehlers, 2008; Peeters, Wessel, Merckelbach, & Boon-Vermeeren, 2002; for a review, see Sumner et al., 2010).

Memory and language: Text-mining and machine-learning approaches

The assessment of AM largely relies on participants’ subjective narratives. Therefore, the linguistic analysis of written AMs could be a useful tool to obtain further insights into the aspects of the architecture of AM that may have psychological or psychopathological relevance. For example, Park, St-Laurent, McAndrews, and Moscovitch (2011) focused on present tense while referring to a past action (i.e., historical present) in autobiographical narratives, suggesting that patients with unilateral temporal lobe excisions used fewer instances of the historical present. This result was interpreted as an indication of a reduced sense of reliving one’s AMs. Other psychoneurolinguistic studies have shown an altered reported-speech use (i.e., direct quotes or indirect paraphrases of someone’s thoughts and words) and reduced creative use of language (i.e., verbal play) in amnesia, which is regarded as evidence of the interdependence across language, memory, and brain functioning (Duff, Hengst, Tranel, & Cohen, 2007, 2009).

Recent technological development in natural language processing (e.g., Nadkarni, Ohno-Machado, & Chapman, 2011) has enabled researchers to analyze respondents’ narratives automatically and comprehensively in order to find specific linguistic features that are relevant to psychopathology, such as

depression and suicidal ideation (e.g., Rosenbach & Renneberg, 2015; Rude, Gortner, & Pennebaker, 2004; Stirman & Pennebaker, 2001). Furthermore, studies combining text-mining and machine-learning approaches have started to develop computerized algorithms that distinguish between writings of people with and without psychopathological problems, for example, genuine versus elicited suicide notes (Pestian, Nasrallah, Matykiewicz, Bennett, & Leenaars, 2010) and tweets of depressed versus nondepressed individuals (De Choudhury, Counts, & Horvitz, 2013). Machine-learning (e.g., Mitchell, 1997) of text categorization is a statistical approach to generate a rule of text classification by learning associations between characteristics of language use (e.g., word choice, sentence length) and specific categories of documents. As Sebastiani (2001) argued, the algorithm generated by such a machine-learning approach often achieves accuracy comparable to that of human experts. Automatic text classification by machine learning is used in various fields of text analysis, such as scoring student essays (Hastings, Hughes, Magliano, Goldman, & Lawless, 2012), testing readability of texts (Sung et al., 2015), and investigating grammatical development of child language (Hassanali, Liu, Iglesias, Solorio, & Dollaghan, 2014). The attempts to develop a classifier of both suicidal notes and depressed tweets have been successful as well, achieving approximately 80% prediction accuracy (De Choudhury et al. 2013; Pestian et al. 2010). However, to the best of our knowledge, no study has applied such approaches to the clinically and theoretically important phenomenon of reduced specificity of AM.

The present study

The present study aimed to explore the linguistic nature of specific memories generated on the AMT by developing an automatic classifier that distinguishes between specific and nonspecific memories conforming to the traditional dichotomous use of AMT responses in the field (Studies 1 and 2), and one that discriminates between the distinctive types of nonspecific memories in a five-category classification of the AMT (Study 3). By the nature of the written version of the AMT, all memories are recorded as sentences, which human raters then classify as specific or nonspecific ones. However, as most of the information that is needed for the specific/nonspecific categorization is in the actual sentences that respondents write down, analyzing the linguistic nature of the written memories could reveal unique characteristics that make a memory (more or less) specific.

In the present study, we followed standard procedures for text classification using machine-learning techniques (Ikonomakis, Kotsiantis, & Tampakas, 2005; Joachims, 2005; Sebastiani, 2001). These include several preprocessing steps such as (a) tokenization and lemmatization, (b) part-of-speech (POS) tagging, and (c) feature selection. First, we

decomposed all the written memories into the smallest grammatical unit, or morpheme (tokenization and lemmatization). For example, a sentence like “I played the guitar” can be decomposed into “I/play-ed/the/guitar.” Second, each morpheme (often referred to as a “token”) was tagged with an appropriate POS on the basis of its definition and context (noun/verb-suffix/article/noun, for this example). Third, we compared the frequency of the morphemes between observer-rated specific and nonspecific memories in order to extract the linguistic features that are most relevant to “memory specificity” (feature selection). Despite the exploratory nature of the present study, we hypothesized that a specific memory would have a greater number of morphemes. A specific (vs. nonspecific) memory normally contains more detailed descriptions of a particular event, that is, contextual information such as temporal, spatial, and sensory–perceptual details, given that a specific memory is defined as a memory referring to a personal past event that occurred at a particular place and time, and lasted for less than one day (Williams et al., 2007). In contrast, expressions referring to extended durations of time (e.g., *life* and *everyday*) and repeatedness of an event (e.g., *always* and *often*) could be, in turn, a strong indication of a nonspecific memory. Feature selection is also of practical importance for machine learning methods, because the reduction in dimensionality (removing irrelevant features for classification) has a number of advantages such as saving computational resources, shrinking the search space, and most importantly, improving classification accuracy (Ikonomakis et al., 2005).

Next, we developed a text classifier by training a support vector machine (SVM; Cortes & Vapnik, 1995), which is a machine-learning technique that shows high performance for two-group classification problems. Using a set of training samples, whose categories (i.e., teacher signal) are already known, a SVM training algorithm develops a “rule” of input features that best classifies a new sample into one category or the other. The SVM was trained to distinguish between specific and nonspecific categories using a document–term matrix consisting of morpheme frequencies as input.

The SVM performance is often assessed by its prediction accuracy for “unknown” test data samples, which were not used in the SVM training processes. Our performance test was twofold. First, we assessed the performance in a test dataset, which comprised AMs that were recalled in the same way (by using the same cue words) as the AMs in the training dataset (Study 1). Second, the SVM was tested in another test dataset, which comprised AMs that were recalled in response to a cue word set that was different from the training dataset (Study 2). This two-step procedure enabled us to examine the general versatility and robustness of the SVM—the extent to which the SVM can predict AM specificity accurately across different memory contents. This performance test is of theoretical importance, as it could reveal whether memory

specificity is determined by “local” words that are specifically associated with the types of cue words, or by general linguistic features that exist across memories triggered by different (sets of) cue words.

We also tested performances of SVM classifiers that were trained only on specific components of language features, such as sentence length, function words, and content words.¹ By comparing the full SVM trained on all the language features with those restricted models, we can explore which linguistic features specifically contribute to the nature of specific memories, and which part of the structural or semantic information of a written AM is important for the accurate classification of AMT responses.

We started with binary classification between specific and nonspecific memories (Studies 1 and 2), although the AMT has four additional categories for the nonspecific memories: extended memories (referring to an event that lasted for more than one day), categoric memories (a category of events or an event that occurred repeatedly), semantic associates (general semantic information), and omission (failure to make a response). Given that most previous studies that used the AMT utilized the proportion of specific memories as an index of memory functioning (i.e., memory specificity), we focused on developing a classifier that detects specific memories out of the other four types of responses. Indeed, Griffith, Sumner, et al. (2012, p. S22), in their review on the assessment of memory specificity, concluded that in most studies in this field, AMT responses are scored and used dichotomously as specific versus nonspecific categories. In Study 3, we evaluated the full five-class classification of the AMT. It was expected that the similarities and ambiguities among “nonspecific” memories might be a challenge to solve the five-class problem (e.g., “Attending a rock fest,” could be an extended memory, a categoric one, or even a semantic associate, depending on the context).²

Study 1

Method

Participants and procedure Participants ($N = 1,240$; 620 men and 620 women; mean age = 44.6 years, $SD = 15.1$ years) were recruited from a Japanese community population by an

¹ In Japanese, function words are grammatically referred as *dependent words* (Hi-jiritsugo), which are always combined with content words (also referred to as *independent words* “Jiritsugo,” such as nouns, verbs, and adjectives) to add extra meaning to phrases and sentences. Function words include particles (“Joshi”) and auxiliary verbs (“Jodoushi”), which have functions similar to prepositions and suffixes in English.

² For example, it has been argued that a large part of semantic associates actually could or should be regarded and treated as a special class of “categoric” responses (Raes et al., 2007).

online survey company (Macromill Inc., Tokyo, Japan). Because all participants had to complete online registration via the website of the survey company, we can assume that most participants had at least minimal knowledge about computers and (the use of) internet. The survey company invited participants, from their list of potential (more than 1 million) respondents, with a small incentive (coupon for online shopping). Participants received explanations about the aims and protocols of the present study, and provided informed consent at the beginning of the survey. All procedures were approved by the institutional review board of Kansai University.

Participants were asked to complete the AMT via the internet with minimal instructions and without any time constraints to recall memories. The Japanese version of the AMT consists of ten emotional (five positive and five negative) cue words that are matched in terms of valence, frequency, and imageability (Gotoh & Ohta, 2001): *self-assured*, *promising*, *succeeded*, *lucky*, and *content* are the positive cues; *discontent*, *desperate*, *fatigued*, *painful*, and *failed* are the negative cues (Takano, Ueno, Mori, Nishiguchi, & Raes, 2016). In response to each cue word, participants were asked to recall and write down an event that they personally experienced in the past. We employed the shortened and simplified version of the instruction used in Heron et al. (2012), which did not explicitly mention that the memory should be specific (see also Debeer, Hermans, & Raes, 2009; Raes, Hermans, Williams, & Eelen, 2007). This minimal set of instructions has been shown to be more sensitive to variations in memory specificity in nonclinical samples, as compared to the traditional instructions in which respondents are explicitly asked to retrieve specific memories (Griffith et al., 2009; Heron et al., 2012). The authors coded all written memories into five categories before statistical analysis, following the coding rules established in previous studies that used the written version the AMT (Heron et al., 2012). The categories were as follows: *specific memory*, which refers to an event that occurred on a particular day (e.g., Last Sunday, I won 150 euros in the National Lottery); *extended memory*, which refers to an extended period of time (i.e., longer than a day; e.g., When I was married to my first wife); *categoric memory*, which refers to a category of similar or repeated events (e.g., I used to buy lottery tickets every Sunday); *semantic associate*, which is not a memory but a mere association to the cue, like an evaluative description of a person or an object (e.g., I'm a lucky person); and *omission*, in which participants did not write down a response. The hand-scoring of the AMT responses identified 3,531 (28%) specific memories, 1,306 (11%) extended memories, 1,644 (13%) categoric memories, 3,406 (27%) semantic associates, and 2,513 (20%) omissions. In order to simplify the classification problem, and because memory specificity is the most important class in the literature (e.g., Williams & Broadbent, 1986), all AMT responses were binary coded as specific versus nonspecific in the analyses of Studies 1 and 2. The interrater

consistency of manual coding was tested on 200 memories that were independently rated by two raters. Interrater agreement was 91% for the two-class categorization, and kappa = .73 for the five-class categorization. The disagreements were resolved through discussion among four of the authors.

Parsing and morpheme analysis All written memories were decomposed into morphemes (or unigrams) by using a text-mining software (KH coder; Higuchi, 2004, 2015), which integrates the Japanese text parser (ChaSen; Matsumoto, 2000), a database, and tools for multivariate statistical analyses. We used the software to obtain a term–document matrix through tokenization (decomposing a sentence into morphemes), lemmatization (standardizing different inflected forms of a word), and POS tagging.

All written memories were collapsed across ten emotional cues because we confirmed that the specificity of a memory was correlated among different cues and represented best by a unidimensional or unifactorial structure (Griffith, Kleim, Sumner, & Ehlers, 2012; Griffith et al., 2009; Heron et al., 2012; Takano et al., 2016³). Across the total of 12,400 memories, the software extracted 6,729 different types of morphemes. At the same time, POS tags were assigned to each word in each memory. After parsing, we calculated three linguistic statistics for each memory: (a) number of morphemes (i.e., token number), which reflects the length of the written memory; (b) number of types of morphemes (i.e., type number), which indicates the vocabulary used in the memory; and (c) type–token ratio (i.e., number of types of morphemes divided by number of morphemes), which reflects the richness of the vocabulary after controlling for the length of the sentence (e.g., Holmes, 1985). We did not include identifiers of cue words in the input vector because such information (which memory was generated in response to which cue words) did not improve the prediction accuracy of the SVM.

Training of the SVMs We divided the 12,400 memories into training and test datasets. Of these, 8,200 randomly sampled memories were used as the training dataset, on which the SVMs were trained. The remaining 4,200 memories (around one third of the total memories; Statnikov, Aliferis, & Hardin, 2011) were used as the test data, in which the classification accuracy of the SVMs was assessed. This division between

³ In this study, we performed an exploratory factor analysis on binary-coded AM specificity, to identify the factor structure of the ten items of the AMT. In line with previous studies (e.g., Griffith et al., 2009), we found that AM specificity has a clear unifactor structure (the eigenvalues were 7.56, 0.60, and 0.37 for one-, two-, and three-factor solutions, respectively), which means that participants who retrieved a specific memory on one cue word tended to retrieve specific memories on the other cue words. In other words, AM specificity can be defined independently of the types of cue words (e.g., different valences). Thus, AM specificity should preferably be scored as a total sum of all cues, not as a subtotal of negative or positive cues (Griffith et al., 2009).

the training and test datasets was fixed across different SVMs, which allowed us to compare the model performance on a constant test dataset. We selected morphemes that should be important for classifying specific and nonspecific memories only from the training dataset. The selection criterion was the relative differences in term frequency (TF; how many times the word was used in a certain group of texts; e.g., Manning, Raghavan, & Schütze, 2008) between the observer-rated specific and nonspecific memories. This was indexed by chi-squared statistics (see also the individual morpheme analysis in the Results and Discussion section). This type of feature vector is often referred to as a document-term matrix, which describes the frequency of terms (or morphemes) that occur in a corpus, with rows as documents (i.e., AMs) and with columns as types of morphemes. Four different feature vectors were prepared as input signals, which consisted of term frequencies of 50, 100, 200, and 500 morphemes,⁴ respectively. We selected these morphemes, which had (a) the highest chi-squared statistics (largest differences in TF between the specific and nonspecific categories), and (b) enough frequencies (more than four in the training dataset). In addition to the morpheme frequency, each of the four feature vectors included the token number and character count of each AM. Character count was included as an additional measure of sentence length because compound (or hyphenated) words are counted as a single token but have more specific meanings and higher character counts.

We employed a Gaussian-kernel SVM, which was trained using five-fold cross-validation within the training dataset. Cross-validation was used to avoid overfitting of the SVM to a single training dataset. To optimize the SVM, we performed a grid search for cost and gamma parameters. However, we found that the parameter optimization did not improve the SVM performances remarkably. Furthermore, because we needed to compare one SVM to the other in the performance tests, we reported the SVMs trained with default values (cost = 1; gamma = 1/data dimensions). Because our training dataset was not balanced between the two classes, we put greater weight on the “specific” than “nonspecific” category (i.e., weight balance = 2:1) in the SVM training.

As a performance test, the trained SVM was used to predict the outcomes (i.e., memory specificity) of the split test dataset. Data splitting was performed because (a) the generalizability of the trained model is not always guaranteed without a validation in an “unknown” test dataset that is separated from the model training process (cf. Esbensen & Geladi, 2010) and (b) we aimed to compare the SVM performance between two different contexts of AM retrieval (i.e., same vs. different cue words; Studies 1 and 2).

⁴ Although we constructed an input vector of 600+ morphemes, the SVM estimation could not be completed due to zero frequencies of some morphemes in the training dataset.

The SVM performance was evaluated by a receiver operating characteristic (ROC) analysis. In this performance test, *hit rate* (ratio of specific memories correctly identified by the SVM relative to all specific memories by observer rating) and *correct rejection rate* (nonspecific memories correctly predicted by the SVM relative to all observer-rated nonspecific memories) were calculated. ROC curves were depicted by plotting the hit rate against the false alarm rate (=1 – correct rejection) at various thresholds of SVM decision values. Note that decision values are continuous scores, which are transformed into binary outputs (i.e., a specific vs. nonspecific category) via a discriminant (or sign) function. Each possible threshold of this discriminant function provides a single point of the ROC curve (e.g., Rakotomamonjy, 2004).

The area under the curve (AUC) was also computed by summing up the areas under each possible point of the ROC curve, which indicates how well a classifier can distinguish between two groups, with .5 indicating a chance level and 1 indicating a perfect separation. Unlike the hit and correct rejection rates, the AUC depicts general behavior of the model that is independent of the thresholds of the SVM decision values. We utilized the R `e1071` package (Meyer, Dimitriadou, Hornik, Weingessel, & Leisch, 2014) for training SVMs, and ROCR package for ROC analysis (Sing, Sander, Beerwinkler, & Lengauer, 2005).⁵

Results and discussion

Token and type numbers As a first step, we tested the differences in basic linguistic statistics (token number, type number, and type–token ratio) between observer-rated specific and nonspecific memories. Means and standard deviations for these basic linguistic statistics are presented in Table 1. The results of an analysis of variance (ANOVA) indicated that token and type numbers had significant differences among the five AMT classes, implying that memory responses (specific, extended, and categoric) tend to be written in longer sentences and to contain a richer vocabulary than nonmemory responses (semantic associates and omissions). Omission and semantic associates had greater type–token ratios than the other three types of memory responses, indicating incompleteness of the sentence rather than richness of the vocabulary; omission and semantic associates typically have fewer words, especially fewer function words (e.g., auxiliary verbs and particles), than specific and the other two types of memory responses.

⁵ The data and R codes for SVM training are available from the Open Science Forum repository: https://osf.io/brdzh/?view_only=6bf54cc871224c429318d19a0cbe46a2. A Web application developed by RShiny (Chang, Cheng, Allaire, Xie, & McPherson, 2015), which embeds the SVM trained on 200 morphemes and works on a Web browser for memory classifications, is also distributed here. (Note that Japanese environments are required in order to make use of this application.)

Table 1 Basic linguistic statistics for Studies 1 and 2

	Specific		Extended		Categoric		Semantic		Omission		ANOVA (Welch)	Par η^2	Multiple Comparisons (Tukey Adjustment)
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>			
Study 1	<i>n</i> = 3,531		<i>n</i> = 1,306		<i>n</i> = 1,644		<i>n</i> = 3,406		<i>n</i> = 2,513				
Token number	11.69	7.75	12.54	9.00	11.32	7.71	5.67	5.74	1.58	0.72	$F(4, 4161.7) = 2,966$	0.29	ex > sp, cat > sem > om
Type number	10.60	5.99	11.30	6.86	10.28	6.03	5.35	4.98	1.58	0.72	$F(4, 4192.8) = 3,792$	0.34	ex > sp, cat > sem > om
Type–token ratio	0.94	0.08	0.94	0.08	0.95	0.08	0.98	0.06	1.00	0.00	$F(4, 4118.8) = 920$	0.11	om > sem > sp, ex, cat
Study 2	<i>n</i> = 1,993		<i>n</i> = 1,173		<i>n</i> = 1,326		<i>n</i> = 2,822		<i>n</i> = 1,164				
Token number	11.60	7.46	12.31	8.11	10.83	7.22	4.70	4.95	1.48	0.69	$F(4, 3334.4) = 2,167$	0.30	ex > sp > cat > sem > om
Type number	10.54	5.81	11.11	6.40	9.96	5.81	4.46	4.32	1.48	0.69	$F(4, 3369.7) = 2,679$	0.34	ex > sp > cat > sem > om
Type–token ratio	0.94	0.08	0.94	0.08	0.95	0.07	0.98	0.05	1.00	0.00	$F(3, 3061.2) = 212^a$	0.11	sem > cat > sp, ex

^a An ANOVA was performed for the four categories other than omissions, which had a constant type–token ratio

Individual morpheme analysis To reveal predictive morphemes that can discriminate between specific and nonspecific memories, we calculated the chi-squared statistics for each morpheme, to determine the extent to which the TFs differed between specific and nonspecific memories. The morphemes were sorted and ranked by the magnitude of the chi-squared statistics in Table 2 (higher chi-squared scores correspond to greater differences in TFs between the specific and nonspecific memory category). This morpheme analysis was performed on the training dataset.

Past-tense auxiliary verbs⁶ (“ta” and “da”) had the largest chi-squared values. On average, past-tense expressions were used more than once in a specific memory, whereas their frequency was only 0.21 in nonspecific memories. Following these past-tense expressions, negation (*no* and *not*) had a greater frequency in nonspecific than in specific memories, because negative expressions were often associated with an omission type of response (e.g., “nothing” and “cannot remember”) and with semantic associates (e.g., “my salary is not enough”). In many cases, expressing the nonexistence of something does not refer to specific experiences, which should include the details of actions, events, and objects that actually took place or existed in the past. For example, a memory “I attended my friend’s birthday party” would be a specific memory, whereas a memory “I did not attend his birthday party” (or “I have never attended his birthday party”) would be scored as a semantic associate unless alternative actions (“because I had to attend a funeral”) are indicated. Furthermore, negation had a strong collocation with *particular* or *particularly*, in the forms of “nothing in particular” and “not particularly.” Because this collocation often appeared in the omission category in the present corpus, the term *particular*(ly) had a relatively large chi-squared value.

⁶ In Japanese, these past-tense expressions are grammatically categorized as auxiliary verbs. Similar to *-ed* in English, “ta” and “da” must be combined with verbs (e.g., “play + ed” in English; “aso-bu + da” in Japanese), although both “ta” and “da” can be conjugated.

Furthermore, some particles (e.g., “de,” “wo,” “ni”) had relatively large chi-squared statistics (Table 2; see also Table A1 in the Appendix), indicating greater frequency of use in specific memories. In Japanese, a particle must be combined with nouns to form an adjectival or adverbial phrase, which provides additional information to a sentence. For example, the particles “de” and “ni” have a function similar to English prepositions (e.g., *at*, *on*, *in*, and *to*) that define the time, direction, and target of an action, when combined with a noun like “I worked *on February 10th*” and “I went *to the park*.” This means that memory sentences including a greater number of adverbial phrases by particles (or prepositions in English) are more likely to be judged as specific, probably because those sentences often have a richer amount of information regarding temporal and spatial details.

Figure 1 shows the proportions of major POS tags (verbs, nouns, adjectives, adverbs, adjectival nouns,⁷ auxiliary verbs, and particles) in the four input vectors of 50, 100, 200, and 500 morphemes. The vector of 50 morphemes consisted of 38 (76%) independent words, which have meaning by themselves (i.e., verbs, nouns, adverbs, adjectival nouns, and adjectives), and 12 (24%) function words, which must be associated with independent words to impart meaning (i.e., auxiliary verbs and particles). This smallest input vector seems to have a minimum vocabulary (see also Table 1) with frequently used independent words (e.g., *do* verb), function words (e.g., auxiliary verbs and particles), and AMT cue words (e.g., *failure* and *success*). A larger input vector had a greater proportion of independent words but a smaller proportion of function words, which implies that the input vector of 500 morphemes covered the richest vocabulary containing less frequently used words that convey specific and unique meanings, as compared to the words in the other input vectors. However, as the number of types of function words is limited

⁷ Adjectival nouns represent a Japanese-specific part of speech, referring to a noun that is given an adjectival function by connecting with the particle “na.”

Table 2 Morphemes that had the largest difference in frequency (indexed by chi-squared statistics) between specific and nonspecific memories

Rank	Morpheme in English	Morpheme in Japanese	Part of Speech	Term Frequency per Memory		Chi-Squared
				Nonspecific	Specific	
1	“ta”/“da” (past tense)	Ta/da	Auxi V	0.21	1.14	1,233.90
2	No	Nai	Adj	0.30	0.02	1,191.93
3	Particularly	Tokuni	Adv	0.11	0.00	452.81
4	Do	Suru	Verb	0.16	0.48	140.04
5	“ha”/“wa” (topic marker, subjective)	Ha	Particle	0.13	0.09	139.83
6	“de” (indicating place, time, method)	De	Particle	0.09	0.29	134.13
7	Be	Aru	Verb	0.07	0.04	110.12
8	Not	Nai	Aux V	0.11	0.09	107.39
9	Living	Seikatsu	Noun	0.03	0.00	94.13
10	“wo” (direct object)	Wo	Particle	0.16	0.44	92.06

Auxi V = Auxiliary verb; Adj = Adjective; Adj N = Adjectival noun; Adv = Adverb. Morphemes were selected from the training data of Study 1 ($N = 8,200$)

by the nature of the language, auxiliary verbs and particles became less dominant in larger input vectors.

Performance test of the support vector machines We trained four different SVMs using different numbers of morphemes (50, 100, 200, and 500) in the document–term matrices with sentence length information (token and character numbers) as input vectors, and the observer-assigned ratings (specific vs. nonspecific memory) as the categorical variable. Each SVM was trained on the training datasets. Then, the SVM performances were tested on the test dataset. The results of the performance test are shown in Table 3. Each model showed good classification performance in the test data, with 86%–87% correct rejections, 83%–87% hits, and .91–.92 AUC values. These results suggest that our SVMs, even the one trained on the smallest input vector of 50 morphemes and sentence length information, showed good performance in classifying the written AMs into specific and nonspecific categories. However, as the number of input morphemes was increased, the model performance improved slightly, although there was no change (or a decrease) in model performance between 50 and 100

morphemes. This small improvement implies that for accurate classification, some memories (around 1% of the total memory samples, in terms of the AUC) require additional knowledge and vocabulary about less frequently used words, which were not included in the base 50 input morphemes.

Although the AUC indicates the SVM with 500 input morphemes as the best model, this model had the lowest hit rate among the four SVMs. This reduction in hit rate implies that the model with 500 morphemes might have overfit the training data, because the input vector could include morphemes that were locally optimal to the training dataset. A similar performance reduction was also found in Study 2. Thus, we regarded the SVM trained on 200 input morphemes as the best model in terms of classification accuracy, and accordingly, this SVM was used for the further analyses. The performance of the SVM on 200 morphemes was also tested for each cue word, and the results ranged from 77.3% to 93.8% hit rates, from 78.6% to 93.0% correct rejection rates, and from .87 to .96 in AUC (see Table 4).

We also examined what linguistic features contributed to the classification accuracy by training SVMs on restricted

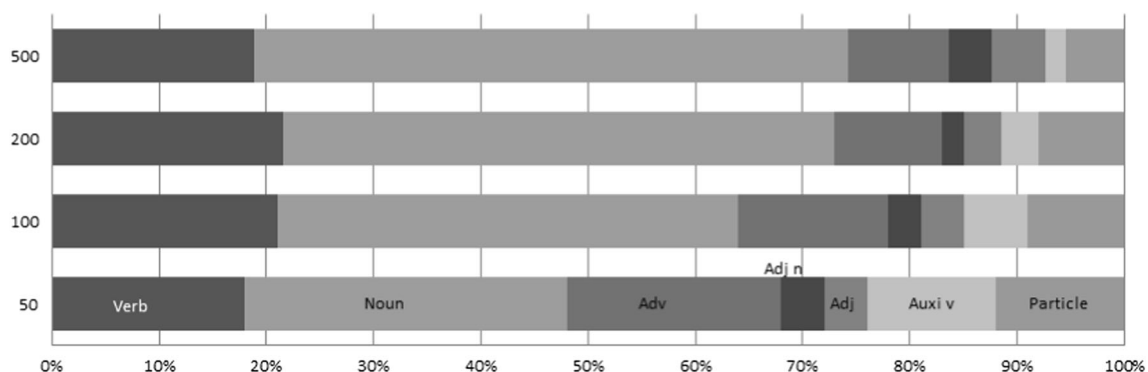


Fig. 1 Proportions of major part-of-speech tags in the four input vectors of 50, 100, 200, and 500 morphemes. Adv = Adverb, Adj n = Adjectival noun, Aux v = Auxiliary verb

Table 3 Results of receiver operating characteristic analysis in Study 1 ($N = 4,200$)

	Hit Rate	Correct Rejection Rate	AUC
Full Models (N morphemes)			
50	.837	.859	.910
100	.834	.859	.909
200	.871	.859	.917
500	.832	.868	.918
Restricted Models			
Sentence length	.898	.541	.751
Content words	.675	.847	.833
Function words	.841	.850	.880

AUC = area under the curve. For restricted models, input vectors were created from sentence length (word and character number), content words (verbs, nouns, adjectives, adjectival nouns, and adverbs), and function words (auxiliary verbs and particles). Content and function words were selected from the 200 morphemes

information from the 200 morphemes: (a) sentence length (character and word counts), (b) 177 content (independent) words (verbs, nouns, adjectives, adjectival nouns, and adverbs), and (c) 23 function (dependent) words (auxiliary verbs and particles). The performance of these three SVMs (see Table 3) suggests that (a) sentence length alone had the smallest signal for the classification of memory responses (AUC = .751), (b) function words alone classified responses better than chance and provided the strongest signal for classification of responses (AUC = .880), (c) content words alone classified responses better than sentence length alone, but by a smaller margin than function words (AUC = .833), and (d) content words provided a classification signal beyond what is provided by function words alone, since the model including all words (AUC = .917) outperformed the model using only function words.

Table 4 Results of receiver operating characteristic analysis for each cue word in Study 1

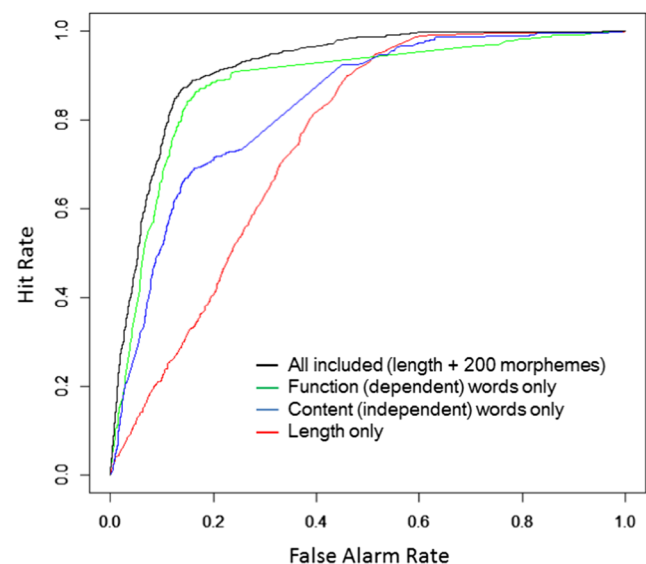
Cue words	N	Hit Rate	Correct Rejection Rate	AUC
Self-assured	420	.938	.873	.944
Discontent	415	.773	.919	.918
Promising	425	.826	.930	.960
Desperate	449	.883	.835	.902
Succeeded	410	.896	.824	.911
Fatigued	423	.827	.821	.872
Lucky	411	.913	.844	.912
Painful	397	.805	.868	.893
Content	418	.891	.836	.921
Failed	432	.840	.786	.870

AUC = area under the curve

Figure 2 shows the ROC curves for the full and restricted SVMs. The full SVM on 200 morphemes still had the best performance over the three restricted models, which suggests the importance of integrating both functional and semantic information in the input feature vector. However, because the function words have a greater contribution to the prediction accuracy than the content words do, binary classification of specific versus nonspecific memories might be more dependent on the structural than on the semantic information of AMs.

Study 2

The results of Study 1 suggest that the SVM trained on 200 input morphemes has sufficient accuracy in classifying AMs into specific and nonspecific memories. One important limitation that should be considered is that the SVM was trained and tested within a single context in Study 1; that is, both training and test datasets comprised of AMs that were retrieved in response to the same ten cue words. Therefore, it is still unclear if the SVM that we developed would also show sufficient accuracy in classifying AMs that were retrieved in response to a different cue word set. Testing the robustness of the SVM in classifying memories that were generated using different (sets of) cue words is especially important given that cue words used in the AMT literature are often different across studies (for a review, see Griffith, Sumner, et al., 2012). Furthermore, clinical interventions such as Memory Specificity Training (MeST; Neshat-Doost et al., 2013; Raes, Williams, & Hermans, 2009), developed to increase the

**Fig. 2** Receiver operating characteristic curves for the support vector machines trained on input vectors of sentence length (red), 177 content words (blue), 23 function words (green), and all of the information (sentence length and 200 morphemes; black)

specificity with which (depressed) individuals retrieve personal memories, use a wider range of cue words than a single AMT assessment.

At the same time, the robustness test of the SVM is of theoretical importance because high performance on this test implies that the SVM captures universal linguistic features of “specificity” that exist across different cue- or trigger-independent memory content. Otherwise, if the SVM performances were substantially worse than in Study 1, this could be an indication of content dependency (or cue word dependency) of memory specificity judgment. By testing the general versatility of the SVM, we could reveal whether memory specificity is coded by individual words and specific vocabulary, or whether it is more dependent on structural information of a sentence such as tense and functional words. Thus, in Study 2, we tested whether the SVM developed in Study 1 holds sufficient accuracy in classifying newly collected AMs that were retrieved in response to cue words different from those used in Study 1.

Method

Participants and procedures Written AMs were collected through an online survey under the same instructions of the AMT in Study 1. Participants ($N = 314$) were recruited by a survey company (Macromill Inc., Tokyo, Japan) from a Japanese community population with balanced age ($M = 44.7$ years, $SD = 18.9$ years) and gender (males = 154, females = 160). Instead of the ten cue words used in Study 1, new cue words (no overlap with Study 1) were used in accordance with previous memory studies and emotional word lists (e.g., Gotoh & Ohta, 2001; Raes et al., 2009). Among the 27 cue words, seven were nouns (i.e., *bicycle, animal, car, gift, party, house, and restaurant*), and the other 20 were adjectives (i.e., *satisfied, healthy, proud, friendly, active, effortful, cheerful, favorable, self-possessed, irritated, assertive, passive, clumsy, handy, happy, sad, hopeful, disappointed, secure, and insecure*). Although the previous memory study (Raes et al., 2009) used more than 50 cue words, we halved the number of the cues to reduce burden on participants. Each participant was asked to write down their memories in response to all the cues. Manual coding of memory specificity was completed by two of the authors (interrater consistency between the two raters on 200 memories was 95% for the binary coding). The manual coding identified 1,993 (23.5%) specific, 1,173 (13.8%) extended, 1,326 (15.6%) categoric memories, 2,822 (33.2%) semantic associates, and 1,164 (13.7%) omission type memories. We confirmed the unifactor structure of the 27 cue words in terms of the specificity (binary coded with 1 as specific and 0 as nonspecific memories). An exploratory factor analysis on the tetrachoric correlations among the cue words revealed eigenvalues of 11.00, 1.95, 1.68, and 1.45 for one- to four-factor solutions, respectively.

All the cues had medium-to-large factor loadings on a single factor ($>.45$) except for the cue words, *satisfied* (.40) and *effortful* (.31). Two thirds of the AMs were randomly selected as training data ($n = 5,600$), and the rest were used as test data ($n = 2,878$). In the training dataset, SVMs were trained as benchmark models that were optimally trained within the memory corpus of the new cue words. Their performances on the training dataset were compared with that of the SVM of Study 1.

Language analysis was identical to Study 1, which utilized the KH coder for Japanese morpheme analysis (Higuchi, 2004, 2015), R e1071 package for SVM development (Meyer et al., 2014), and R ROCR package for ROC analysis (Sing et al., 2005). Morphemes that should be used to develop benchmark SVMs were selected from the training dataset by using chi-squared statistics as an index of TF differences between specific and nonspecific AMs. In addition, we excluded morphemes with low frequencies (less than 5 in the training dataset of Study 2) to avoid constant predictors in a cross-validation routine of SVM training.

After this feature selection, benchmark SVMs were trained on three different sets of input morphemes (i.e., 100, 200, and 500). Model performances were tested on the test dataset, which was not used in the SVM training process. All the SVMs were trained with the Gaussian kernel, five-fold cross-validation, and class weights (i.e., 3:1 for specific and nonspecific categories).

Results and discussion

Among the total 8,476 memory responses, 1,993 were rated as “specific” and the other 6,483 memories were rated as “nonspecific,” by manual coding. Similar to the results of Study 1, the basic linguistic statistics (Table 1) indicate that a specific memory is characterized by a greater number of words (token number) and a richer vocabulary (type number). Table 5 shows the results of the performance test of (a) the SVMs trained on the training dataset of Study 2 and (b) the SVMs imported from Study 1. Among the SVMs that were trained within the dataset of Study 2, the SVM of 200 morphemes exhibited the highest performance in predicting manually coded memory specificity, achieving .92 AUC. Although the model trained in Study 1 (for 200 input morphemes) had a slightly worse performance than that of Study 2, it still performed well ($AUC = .888$), regardless of the cue word differences. However, the SVM trained in Study 1 showed relatively variable performances for individual cue words, ranging from 55.6% to 95.0% in hit rate, and 57.5% to 95.1% in correct rejection rate, and .78 to .98 in AUC (Table 6). The worst performance was observed for the cue words *satisfied* and *effortful*, which had the lowest factor loadings in the exploratory factor analysis. Responses to these cue words were different from those to the other cue words,

Table 5 Results of the receiver operating characteristic analysis in Study 2 ($N = 2,878$)

	Hit Rate	Correct Rejection Rate	AUC
SVMs trained on the training dataset of Study 2: (N morphemes)			
100	.796	.875	.914
200	.831	.871	.919
500	.730	.892	.918
SVMs trained on the training dataset of Study 1:			
Full model (200 morphemes + sentence length)	.830	.823	.888
Restricted model (content words only)	.640	.784	.765
Restricted model (function words only)	.796	.847	.874

AUC = area under the curve. The full model included all the information about sentence length, content words (177 tokens), and function words (23 tokens)

Table 6 Results of receiver operating characteristic analysis for each cue word in Study 2

	N	Hit Rate	Correct Rejection Rate	AUC
Satisfied	108	.824	.622	.781
Bicycle	108	.950	.588	.874
Healthy	102	.750	.802	.836
Animal	111	.815	.833	.902
Proud	110	.818	.869	.877
Car	106	.839	.773	.871
Friendly	104	.852	.883	.919
Gift	118	.878	.727	.862
Active	100	.808	.730	.838
Party	113	.825	.768	.836
Effortful	103	.813	.575	.811
House	94	.900	.844	.946
Cheerful	95	.833	.888	.953
Restaurant	100	.909	.791	.911
Favorable	121	.833	.861	.935
Self-possessed	116	.929	.775	.921
Irritated	86	.708	.887	.847
Assertive	114	.571	.880	.825
Passive	112	.750	.837	.915
Clumsy	112	.786	.888	.890
Handy	91	.900	.951	.975
Happy	113	.743	.846	.870
Sad	110	.814	.940	.916
Hopeful	112	.889	.913	.977
Disappointed	108	.857	.894	.935
Secure	104	.851	.842	.906
Insecure	107	.556	.908	.893

AUC = area under the curve

which might have influenced the prediction accuracy of the SVM. The hit rate was particularly low for the cue words *assertive* (57.1%) and *insecure* (55.6%). However, because the proportions of specific memories were relatively small for these two cue words (12% and 8%), as compared to the overall average (23.5%), sampling biases as well as content differences may have influenced the SVM performance.

With reference to the performance of the restricted models, the SVM trained on content words showed worse performance in Study 2 than in Study 1 (AUCs = .765 and .833, respectively). On the other hand, the SVM trained on function words exhibited similar levels of performance between Studies 1 and 2 (AUC = .880 and .874, respectively). These results suggest that (a) the versatility of the full SVM on 200 morphemes could be mainly attributed to function words, and (b) content words should be selected according to the set of cue words to improve the prediction accuracy of the SVM.

Among the 200 input morphemes that were used to train the SVMs in Studies 1 and 2, 72 morphemes (17 function words) were commonly used between the two SVMs (a full list of these morphemes has been provided in the [Appendix](#)). These common morphemes reflect general features of specific memories: past-tense auxiliaries, prepositions (or particles in Japanese), negations (*no* and *not*), and temporal expressions (e.g., *life*, *always*, and *everyday*), which exist across diverse contents of memories generated on different cue words. The other 128 morphemes, which were newly selected following the feature selection in Study 2, slightly improved the classification accuracy of the SVM developed in Study 2. This result suggests that for some memories, extra vocabulary that is specifically associated with the cue words contributes to enhanced classification accuracy; for example, the phrase “In a wedding ceremony” in a memory cued by the word *party* is a clear indication of specificity, because human raters have a priori knowledge that a wedding ceremony normally takes place within a single day.

Study 3

Studies 1 and 2 demonstrated the performance of the SVM in the binary discrimination between specific and nonspecific memories. To evaluate our machine-learning approach in a five-class classification of the AMT, we repeated similar analyses using the same datasets, but employing the multiclass SVM algorithm in Study 3.

Method

Feature selection To select the morphemes that can be used in multiclass SVM training, we first computed chi-squared statistics for each morpheme by a one-versus-all comparison (e.g., specific memories vs. the other four categories) in the whole dataset of Study 1. Next, we selected potentially predictive morphemes by using the Round-Robin algorithm (i.e., selecting an equal number of candidates from each of multiple categories; Forman, 2004, 2007). The one-versus-all comparison was repeated for each of the five AMT categories, and in each comparison loop (extended vs. the other four, categoric vs. the other four, etc.), the same number of morphemes were nominated on the basis of the chi-squared statistics. For each AMT category, this nomination process revealed 186 morphemes that had the largest chi-squared statistics (>3.0) and sufficiently high frequency for analysis (>4). One exception was the *Omission* category, which identified only 78 available morphemes due to the limited vocabulary. This feature selection resulted in 514 morphemes after exclusion of overlaps in the nomination across multiple classes. Among the 514 morphemes, 192 morphemes were also used in the binary SVM trained on 200 morphemes that was used in Study 1.

SVM training We found that the unbalanced sample sizes across the five AMT categories caused a serious bias in SVM training; therefore, we prepared a balanced dataset through sampling 1,000 responses randomly from each AMT category except for extended memories (840 responses) as a training dataset (4,840 responses in total), and 300 from each category as a test dataset (1,500 in total). These sample sizes were determined because the dataset of Study 1 had 1,306 extended memories (840 in the training and 466 in the testing data after random data splitting), which was the smallest class across the five AMT categories. To test the general versatility of the five-class SVM across different sets of cue words, we tested the performance of the SVM in predicting the data collected in Study 2. For ease of interpretation and comparison between the two datasets, we sampled 300 responses randomly from each of the five classes in the dataset of Study 2. Because the test data were balanced across 5 classes, we used accuracy (all correct predictions of specific and nonspecific memories relative to total samples) as a performance index in Study 3.

However, as an SVM can solve only a binary problem, the five-class SVM was trained using the “one-against-one” approach. In this training algorithm, binary classifiers were trained for each pair of classes, and each binary classifier “votes” to determine appropriate classes of test samples (Meyer et al., 2014). In this voting system, ten binary classifiers that were trained to distinguish between two given categories for ten possible combinations (e.g., specific vs. extended, specific vs. categoric, etc.) determined the most appropriate class for each memory response by a majority rule. Other settings were the same as the SVM training in Studies 1 and 2; we used the R e1071 package (Meyer et al., 2014), and the SVM was trained with the Gaussian kernel and five-fold cross-validation.

Results and discussion

The five-class SVM achieved 65.1% prediction accuracy as an average of five-fold cross-validation in the training dataset. Given that the chance level is 20% under the five-class problem, this accuracy means that the SVM has better separation than a random classification. In predicting the test dataset, the SVM classifier showed 64.7% accuracy (see Table 7). This performance was retained in predicting the test dataset from Study 2, which achieved an accuracy of 63.7%. However, the sensitivity to detect extended memories dropped from 47.3% in Study 1 to 42.7% in Study 2. This reduction in sensitivity suggests that the identification of extended memories needs more cue-dependent information and a priori knowledge about the tokens (see the “wedding ceremony” example in

Table 7 Confusion matrix of predicted and actual categories by the five-class support vector machine (SVM)

SVM prediction	Observer Rating				
	Specific	Extended	Categoric	Semantic	Omission
Study 1 test dataset					
Specific	186	67	38	14	0
Extended	54	142	47	41	0
Categoric	44	56	161	42	0
Semantic	15	35	53	197	15
Omission	1	0	1	6	285
Study 2 test dataset					
Specific	174	51	34	22	0
Extended	53	128	31	19	0
Categoric	51	81	167	54	0
Semantic	22	40	67	191	5
Omission	0	0	1	14	295

The five-class SVM was trained on the training dataset of Study 1 ($n = 4,840$), and applied to the test datasets of Studies 1 and 2. The test datasets were created through random sampling of 300 responses for each category

Study 2) than the identification of other types of memories. Importantly, the patterns of classification errors appear to reflect the similarities between the types of memories. For example, specific memories were often misclassified as extended memories, and vice versa. Furthermore, categoric memories and semantic associates tend to be confused for each other, which suggests that categoric memories share some language features that can be seen in nonmemory responses, and vice versa. The latter finding is in support of an earlier claim by Raes et al. (2007, p. 496) that a large part of semantic associates actually could or should be regarded and treated as a special class of “overgeneral categoric” responses.

Overall, the performance of the five-class SVM classifier was better than chance, but the prediction accuracy was still moderate. A possible reason for the lower performance of the five-class SVM relative to the binary classifier may be that misclassifications often occurred among the three types of “nonspecific” responses (i.e., extended, categoric, the semantic associates), which accounted for more than half of the total classification errors (51.8%). Furthermore, the proportion of the input features that were specifically selected to detect specific memories was relatively small in the training dataset of the five-class model, because we needed to add extra features that are associated with categories other than specific memories. This proportional reduction of “specific” features would influence the sensitivity to detect specific memories, which was decreased in the five-class SVM (62.0%), as compared to that of the binary model (around 80%).

General discussion

In the present study, we sought (1) to reveal the linguistic features of specific memories on the AMT and (2) to develop an automatic categorization algorithm of memory specificity by employing text-mining and machine-learning approaches. From the morpheme analysis, we learned that specific memories tend to contain a greater number of words than do nonspecific memories, and that some types of morphemes (e.g., past-tense auxiliary verbs) are specifically associated with specific memories. The SVM classifier that we developed showed good performance, with an AUC of .89–.92 in categorizing specific and nonspecific memories of “unknown” test datasets (Studies 1 and 2), and with 64%–65% accuracy in the full five-class categorization (Study 3). These results demonstrate the feasibility of categorizing AMT responses as specific versus nonspecific by using a computerized classification algorithm, although the five-class classification using more fine-grained distinctions between categories of memories is more difficult than the binary classification.

Our results have four main implications regarding the linguistic nature of specific AMs. First, most specific memories are written in past tense. Past-tense expressions by auxiliary

verbs “ta” and “da” in Japanese (cf. *-ed* in English) were observed more than once in specific memories, whereas nonspecific memories contained the past-tense expressions only 0.21 times. Given that a memory represents a past event, these results are quite evident and straightforward. In other words, this finding suggests that a substantial number of “memory” responses had no grammatical indications of past tense. Such responses could be categoric memories, semantic associates, or omissions, which are often described in the present tense (“whenever other people disappoint me,” categoric memory; “I’m a lucky person,” semantic associate) or in an incomplete one-word sentence (“nothing,” omission).

Second, our results suggest that memory responses (specific, extended, and categoric memories) contain a greater amount of information than do nonmemory responses (semantic associates and omissions), as memories are written in longer sentences (i.e., token number) with a richer vocabulary (i.e., type number). Although sentence length can be a filter for screening out semantic associates and omissions, the model trained only on sentence length had lower performance when differentiating a specific memory from the other types of categories, compared with models trained on content and/or function words.

Related to this point, adverbial phrases led by prepositions (particles in Japanese) are important indicators of memory specificity. Some prepositions, such as *to* and *in* (“ni” and “de” in Japanese), add detailed information about time and location to sentences. This then increases the probability of classifying a written memory as specific. For example, a simple sentence such as “I went” could be described in more detail by adding prepositional phrases, “I went *to* the park *on* February 10th,” which is more likely to be judged as specific. Thus, including the information on when, where, what, and other modifiers would be a critical factor for the specificity judgment, although the specific objects governed by the prepositions (or particles) are not necessarily important.

Third, the results of the performance tests suggest that most of the information needed for accurate judgment of memory specificity is already included in the smallest input morpheme vector, as the SVM trained on the input vector of 50 morphemes showed an AUC of .91. Although the prediction performances were improved by extending the feature vectors to 100 and 200 morphemes, memory specificity of 80%–90% memories is correctly judged by the usage pattern of 50 basic morphemes, consisting of 12 function words (e.g., auxiliaries and particles) and 38 independent words (e.g., adjectives and nouns).

Fourth, the results of the restricted models suggest that the function words are more predictive of AM specificity than are content words. The SVM trained only on content words showed worse performance in Study 2 than in Study 1, whereas the SVM trained only on function words exhibited similar levels of performance between Studies 1 and 2. These results suggest that the versatility of the full SVM on 200 morphemes can be mainly attributed to function words. In other words,

memory specificity is more likely to be determined by functional and configurational linguistic features such as verb tense, negation, and the number of particles (how the memory was described) than by the individual meanings of words and phrases (i.e., the actual content of the memory).

Among the content (or independent) words were several time-related words, such as *today* and *everyday*, which are predictive of memory specificity. According to the standard AMT coding procedure (Williams & Broadbent, 1986), a specific memory should refer to an event that happened on one particular day. In line with this coding criterion, the words *today*, *the other day*, and *yesterday* were more frequently observed in specific than in nonspecific memories, whereas *everyday*, *life*, *daily life*, and *days* were more often observed in nonspecific (extended and categoric) than in specific memories. These content words nevertheless add important (although limited) information over the function words for the classification of specific and nonspecific memories, because the classification accuracies of the full SVM and content-word SVM are lower than those of the SVMs optimally trained in the dataset of Study 2. This performance reduction means that for some memories, additional vocabulary and a priori knowledge about the words (e.g., a *wedding ceremony* normally does not last longer than a single day) need to be installed in the model. This point is particularly important for the five-class classification in Study 3, as the sensitivity to detect extended memories considerably differed between the two test datasets.

It is also noteworthy that the morphemes that correspond to the cue words (*failure*, *success*, *contentment*, *despair*, *lucky*, *discontentment*, and *self-confidence*) are listed as the most informative words to distinguish between specific and nonspecific AMs (see Table 2 and the Appendix). Some cue words (*failed* and *succeeded*) were more likely to evoke specific memories, whereas others (*discontent*, *self-assured*, *promising*, and *painful*) were more strongly connected to nonspecific memories or unsuccessful retrieval of AMs. Previous studies have already revealed such variability in sensitivity and discriminatory power among different cue words, using item response theory (Griffith, Kleim, et al., 2012; Griffith et al., 2009; Heron et al., 2012). Thus, importantly, SVM learning would be influenced by the set of cue words. Therefore, to some extent, our trained SVM was localized to the cue words that we used in the present study, although it should be noted that the SVM did show a high level of versatility across different cue word sets.

From these analyses, it is possible to speculate about the memory-language architecture underlying specific AM retrieval. The influential memory model of Conway and Pleydell-Pearce (2000) suggests that overgeneral memory is the result of a truncation (or dysfacilitation) of a top-down memory search that is “stuck” at too high a level of a hierarchical memory structure (Williams et al., 2007). In such a case, the generation phase of retrieval is terminated before a detailed memory representation is formed, and thus, only the

general descriptive information can be accessed (Conway & Pleydell-Pearce, 2000; Williams et al., 2007). Consistent with this theoretical account, our observations suggest that a non-specific memory has a shorter sentence length and fewer adverbial phrases led by particles (or prepositions) than a specific memory does. The lack of adjectival and adverbial expressions about experiential, temporal, and spatial details could reflect the hypothesized truncation of memory search. However, another possibility remains—namely, that a failure in verbalization, rather than a truncation of retrieval per se, is associated with the less detailed description of a “nonspecific” memory. In the AMT, and particularly in the written version, participants can easily choose to suppress the verbal report of experiential details, even though they successfully and vividly retrieved such information from memory.

When interpreting our results, the following limitations should be borne in mind. First, our analysis is largely dependent on the Japanese language structure. Although we have already confirmed that the same text-mining and machine-learning approach works well for Dutch written memories (Takano, Ueno, Mori, Nishiguchi, & Raes, 2015), future research needs to test whether similar results can be obtained for other languages. Second, computer-based morpheme decomposition is not always accurate. Particularly in Japanese, words are not separated by a space, unlike European languages, which sometimes causes errors in detecting individual morphemes. Although the software used in the present study is highly reliable when it comes to analyzing formal texts (Higuchi, 2004), it was not able to cover all types of compounds and slangs. Third, there is some ambiguity in the original coding rules of the AMT. Particularly in the written version of the AMT, most memory responses have only one short sentence (sometimes two). Furthermore, as some memories lack information for accurate coding, human raters often have to use their own judgment. This ambiguity results in discrepancies in coding between two independent raters. Although the interrater reliability normally lies around a 90% concordance rate ($\kappa = .80$) in the written version of the AMT (e.g., Heron et al., 2012; Raes, Hermans, de Decker, Eelen, & Williams, 2003), around 10% of the memories are difficult to categorize even by manual coding. Thus, we speculate that the technical errors in creating input vectors (morpheme detection) and ambiguity in teacher signals (categories given by the manual coding) might have influenced SVM learning, which, at least partly, contributed to the classification errors observed in our ROC analysis. Fourth, we did not include any questions to screen out participants’ satisficing (e.g., Krosnick, 1991), which is a behavior to provide satisfactory answers without paying enough attention and effort. Although we believe that our language data are qualified by the observer rating, there might have been sampling biases due to such compliance issues (e.g., intentionally shorten a response to complete the survey quickly).

Notwithstanding these limitations, we have demonstrated that the computerized classifiers that have been developed by using

several linguistic features of specific memories achieves a high classification accuracy when categorizing a new memory as “specific” or “nonspecific.” This high performance of the SVM suggests that memory specificity can be accurately judged by the usage pattern of a relatively small set of words (around 200 morphemes), implying that memory specificity is primarily determined by the functional and configurational information of a sentence (e.g., auxiliary verbs and particles), and to a lesser extent by the semantic elements of a written memory. As a practical implication, our SVM could contribute to the automation and standardization of AMT coding protocols. As we mentioned in the introduction, an increasing number of studies are using the AMT, given the clinical and theoretical importance of memory specificity. Although the present studies are limited to the two-class (specific vs. nonspecific) classification, the SVM classifier would save expert manpower and enable researchers to make objective and standardized judgments of AMT data. Furthermore, such automation opens new avenues with reference to going online with clinical interventions directed at the remediation of overgeneral memory, like MeST (Raes et al., 2009).

Author note This study was supported by the Japan Society for the Promotion of Science grant awarded to K.T. (24-1573) and J.M. (25885079). F.R. was supported by the KU Leuven Research Council grant PF/10/005.

Appendix

Table 8 Input morphemes used in Studies 1 and 2 (for training support vector machines on 200 input morphemes)

Morpheme (English)	Morpheme (Japanese)	Part of Speech	Chi-Squared	
			Study 1	Study 2
“ta” and “da” (past tense)	Ta/da	Aux V	1,233.9	828.6
No	Nai	Adj	1,191.9	443.9
Particularly	Tokuni	Adv	452.8	171.1
Do	Suru	Verb	140.0	11.0
“ha” (topic marker, subjective)	Ha	Particle	139.8	101.8
“de” (place, method)	De	Particle	134.1	65.3
Be (for an object)	Aru	Verb	110.1	8.4
Not	Nai	Aux V	107.4	100.7
Living	Seikatsu	Noun	94.1	36.0
“wo” (direct objective)	Wo	Particle	92.1	68.3
Now	Ima	Adv	75.8	18.1
Everyday	Mai-nichi	Adv	73.9	51.5
Passing an exam	Goukaku	Noun	73.5	8.5
Life	Jinsei	Noun	56.6	21.1
“mo” (too, as well)	Mo	Particle	51.8	39.7
Eat	Taberu	Verb	47.3	7.3
Buy	Kau	Verb	45.3	35.0
“ni” (place, time)	Ni	Particle	38.1	31.3
“da” (decisive)	Da	Aux V	37.8	66.3
Days	Hibi	Adv	35.5	12.4
Always	Itsumo	Adv	34.1	24.2
Today	Kyou	Adv	33.2	12.7
First time	Hajimete	Adv	31.8	38.9
Understand	Wakaru	Verb	31.4	5.9

Table 8 (continued)

Morpheme (English)	Morpheme (Japanese)	Part of Speech	Chi-Squared	
			Study 1	Study 2
“aru” (perfect tense)	Aru	Aux V	29.3	8.4
Work	Shigoto	Noun	29.2	22.8
“n” (negation)	N	Aux V	27.4	7.0
Yesterday	Kinou	Adv	25.0	17.1
Come up with	Omoitsuku	Verb	24.8	8.4
“ga” (subjective)	Ga	Particle	23.6	7.4
In the future	Shourai	Adv	22.1	11.8
Relationship	Kankei	Noun	21.3	8.2
The other day	Senjitsu	Adv	21.2	15.5
Go	Iku	Verb	20.9	5.3
Be (for a person)	Iru	Verb	32.9	15.5
Bicycle	Jitensha	Noun	19.4	6.4
Health	Kenkou	Noun	18.4	13.2
Tournament/conference	Taikai	Noun	17.7	20.8
Growth	Seichou	Noun	17.3	9.2
“ya” (enumeration)	Ya	Particle	17.2	17.4
“nado” (exemplification)	Nado	Particle	16.6	11.9
Be accepted	Ukaru	Verb	16.6	5.6
Wonder	Omou	Verb	16.6	9.0
Recently	Saikin	Adv	15.9	8.1
Self	Jibun	Noun	15.5	5.7
Last year	Sakunen	Noun	14.0	18.5
Come to mind	Ukabu	Verb	13.5	8.9
Many	Ooi	Adj	13.1	11.3
Not so much	Amari	Adv	12.9	16.9
Birth	Tanjou	Noun	12.4	82.8
Death	Shi	Noun	12.4	5.6
Body	Karada	Noun	11.6	10.3
Accident	Jiko	Noun	10.4	5.8
Economics	Keizai	Noun	10.3	5.8
“tari” (exemplification)	Tari	Particle	10.3	7.5
Being long time	Hisashiburi	Noun	9.9	5.8
Purchase	Kounyuu	Noun	9.8	31.0
Winning a victory	Yuusyou	Noun	9.7	13.7
Betray	Uragiru	Verb	9.7	5.5
Operation	Syujutu	Noun	9.6	5.6
Be born	Umareru	Verb	8.6	12.6
Against	Ni-taishite	Particle	8.5	5.1
Christmas	Kurisumasu	Noun	8.5	32.8
Telephone	Denwa	Noun	8.3	8.1
Die	Shinu	Verb	8.3	13.8
Future	Mirai	Noun	8.2	8.2
Be decided	Kimaru	Verb	8.1	9.2
Gift	Purezento	Noun	7.4	50.6
Think	Kangaeru	Verb	7.4	5.6
Diagnosis	Shindan	Noun	7.1	16.3
“ba” (subjunctive)	Ba	Aux V	6.8	8.6
“to” (parallel marker)	To	Particle	6.6	12.3
Self-assurance	Jishin	Noun	63.3	–
Discontent	Fuman	Noun	52.8	–
Win (a lottery)	Ataru	Verb	45.2	–
Promising	Yuubou	Adj V	42.1	–
Elected/winning a lottery	Tousen	Noun	32.4	–
Failure	Shippai	Noun	32.1	–
Tasty	Oishii	Adj	31.1	–
University	Daigaku	Noun	31.0	–
Painful	Kutsuu-na	Adj V	29.8	–
“masu” (polite)	Masu	Aux V	29.7	–
Cake	Ke-ki	Noun	27.7	–
Success	Seikou	Noun	26.7	–
Exam	Shiken	Noun	23.7	–
Despair	Zetsubou	Noun	22.6	–
Child rearing	Kosodate	Noun	22.0	–
Live	Ikiru	Verb	21.8	–
Pain	Kutsuu	Noun	19.6	–
Family	Kazoku	Noun	19.4	–
Lottery	Takara-kuji	Noun	19.0	–
Taking an exam (school entrance/skill qualification)	Juken	Noun	18.0	–
Child	Kodomo	Noun	16.9	–

Table 8 (continued)

Morpheme (English)	Morpheme (Japanese)	Part of Speech	Chi-Squared	
			Study 1	Study 2
Human	Ningen	Noun	16.2	–
Little	Sukunai	Adj	15.7	–
Make a mistake	Machigaeru	Verb	15.3	–
Entrance exam	Nyuushi	Noun	15.3	–
Aspiration	Shibou	Noun	15.2	–
Care/Nursing	Kaigo	Noun	14.9	–
All	Subete	Adv	14.9	–
Can remember	Omoi-daseru	Verb	14.9	–
Enter	Hairu	Verb	14.4	–
Currently	Genzai	Adv	14.3	–
Compliment	Homeru	Verb	14.1	–
Fracture (of a bone)	Kossetsu	Noun	14.1	–
House work	Kaji	Noun	14.1	–
Experience	Keiken	Noun	14.1	–
Prize	Kensyou	Noun	14.0	–
Candy	Okashi	Noun	13.8	–
Japan	Nihon	Noun	13.6	–
Feel	Kanjiru	Verb	13.3	–
Because	Node	Particle	13.0	–
Know	Shiru	Verb	13.0	–
Salary	Kyuuryou	Noun	12.9	–
Make	Tsukuru	Verb	12.8	–
Cloth	Fuku	Noun	12.7	–
Getting a job	Syuu-syoku	Noun	12.6	–
Can buy	Kaeru	Verb	12.4	–
Memory	Kioku	Noun	12.4	–
Commodity	Syuu-hin	Noun	12.2	–
Content	Manzoku	Noun	11.7	–
Work	Hataraku	Verb	11.6	–
Tasty (notation variability)	Oishii	Adj	11.4	–
Daily	Nichijou	Noun	11.4	–
Service/pay	Taiguu	Noun	11.4	–
Only	Shika	Particle	11.3	–
Yen (Japanese money)	En	Noun	11.3	–
Being late at	Chikoku	Noun	11.3	–
Informal decision (for employment)	Naitei	Noun	11.3	–
Bus	Basu	Noun	11.3	–
Presentation	Purezzen	Noun	11.1	–
Long	Nagai	Adj	11.0	–
Fully	Ippai	Adv	10.7	–
Shop	Mise	Noun	10.4	–
To/against	Ni-taisuru	Particle	10.3	–
High school	Kou-kou	Noun	10.0	–
Environment	Kankyuu	Noun	9.9	–
Score	Tensuu	Noun	9.9	–
Supermarket	Su-pa-	Noun	9.9	–
Sale	Se-ru	Noun	9.9	–
Carrying/Mobile phone	Keitai	Noun	9.9	–
Successful bid	Rakusatsu	Noun	9.9	–
Shopping	Syoppingu	Noun	9.9	–
Pick up	Hirou	Verb	9.9	–
Blessed	Megumareru	Verb	9.9	–
Drop	Otosu	Verb	9.7	–
Hit	Ataru	Verb	9.7	–
Walk	Aruku	Verb	9.3	–
Money	Kinsen	Noun	9.2	–
Income	Syuunyuu	Noun	9.2	–
Pension	Nenkin	Noun	8.9	–
From now on	Korekara	Adv	8.9	–
Room	Heya	Noun	8.7	–
Interview	Mensetsu	Noun	8.7	–
Train	Densya	Noun	8.6	–
Be able to	Dekiru	Verb	8.6	–
Husband	Otto	Noun	8.6	–
Home	Family	Noun	8.5	–
Horse race	Keiba	Noun	8.5	–
Bonus	Bo-nasu	Noun	8.5	–
Application	Oubo	Noun	8.5	–
Lose (somebody)	Nakusu	Verb	8.5	–

Table 8 (continued)

Morpheme (English)	Morpheme (Japanese)	Part of Speech	Chi-Squared	
			Study 1	Study 2
Picture	Syashin	Noun	8.3	–
Break up	Wakareru	Verb	8.3	–
Ride	Noru	Verb	8.3	–
Say	Iu	Verb	7.8	–
Now	Ima	Adv	7.8	–
Should	Beshi	Aux V	7.8	–
Tasty	Umai	Adj	7.7	–
More	Motto	Adv	7.7	–
Even	Demo	Particle	7.5	–
Tiredness	Tsukare	Noun	7.5	–
Customer	Okyaku	Noun	7.4	–
Quit	Yameru	Verb	7.4	–
Lottery	Kuji	Noun	7.4	–
Take	Ukeru	Verb	7.4	–
Oneself	Jishin	Noun	7.3	–
Work place	Syokuba	Noun	7.2	–
Korea	Kankoku	Noun	7.1	–
(Rock) Band	Bando	Noun	7.1	–
Answer	Kaitou	Noun	7.1	–
Injection	Chuusya	Noun	7.1	–
Winning a prize	Nyuusyuu	Noun	7.1	–
Important	Juuyou	Adj V	7.1	–
Tokyo	Toukyou	Noun	7.1	–
Loss	Son	Noun	7.1	–
Hit/strike	Butsukeru	Verb	7.1	–
Get lost	Mayou	Verb	7.1	–
Lunch	Ranchi	Noun	7.1	–
Lottery	Kuji-biki	Noun	7.1	–
Letter	Hagaki	Noun	7.1	–
Buffet-style	Baikingu	Noun	7.1	–
Suddenness	Totsuzen	Noun	7.0	–
Auction	O-kusyuu	Noun	7.0	–
Reservation	Yoyaku	Noun	7.0	–
Child rearing	Ikuji	Noun	6.8	–
First /top	Ichiban	Adv	6.7	–
Let somebody in	Ireru	Verb	6.7	–
Drink	Nomu	Verb	6.5	–
Years	Nenkan	Noun	6.4	–
“no” (Possession)	No	Particle	–	60.6
Party	Pa-thi-	Noun	–	37.4
Friend	Tomodachi	Noun	–	37.4
Receive	Morau	Verb	–	35.1
“tai” (willingness, would like to)	Tai	Aux V	–	34.0
Clumsy	Bukiyuu	Adj V	–	24.9
Favorable/going smoothly	Junchou	Adj V	–	24.4
Calm/Be self-possessed	Ochitsuku	Verb	–	22.3
Wedding ceremony	Kekkon-shiki	Noun	–	19.9
Become	Naru	Verb	–	19.4
Pass away	Nakunaru	Verb	–	19.1
Being passive	Syoukyoku	Noun	–	19.1
Newly built house	Shin-chiku	Noun	–	19.0
Anxiety	Fuan	Adj V	–	18.8
Effort	Doryoku	Noun	–	17.8
Good at	Tokui	Adj V	–	17.8
Result	Kekka	Noun	–	17.0
Receive (Kanji notation variability)	Morau	Verb	–	16.3
“ka” (interrogation)	Ka	Aux V	–	14.4
Party (notation variability)	Pa-thi	Noun	–	14.3
Until	Made	Particle	–	14.2
Feeling easy	Anshin	Adj N	–	14.0
Personality	Seikaku	Noun	–	13.7
New year	Shin-nen	Noun	–	13.3
Last year	Kyonen	Noun	–	12.9
Being bad at	Nigate	Adj V	–	12.8
Give	Ageru	Verb	–	12.1
Wait	Matsu	Verb	–	11.8

Table 8 (continued)

Morpheme (English)	Morpheme (Japanese)	Part of Speech	Chi-Squared	
			Study 1	Study 2
Safely	Buji	Adv	–	11.6
Forward/before/ex-	Mae	Noun	–	11.5
Mother	Haha	Noun	–	11.4
Celebration	Iwai	Noun	–	11.4
Necklace	Nekkuresu	Noun	–	11.4
Snow	Yuki	Noun	–	11.4
Breaking down	Taoreru	Verb	–	11.4
Desired	Hoshii	Adj	–	11.1
Hope	Kibou	Noun	–	11.0
Like	Suki	Adj V	–	10.7
Being active	Sekkyoku	Noun	–	10.7
Study	Benkyou	Noun	–	10.4
“desu” (politeness)	Desu	Particle	–	10.1
Ceremony (for marriage)	Hirou-en	Noun	–	10.0
Second hand	Chuuko	Noun	–	10.0
Parents	Oya	Noun	–	9.7
Finally	Yatto	Adv	–	9.5
Granma	Baa-chan	Noun	–	9.5
Fall/ fail	Ochiru	Verb	–	9.5
Find	Mitsukaru	Verb	–	9.5
Steal	Nusumu	Verb	–	9.5
Cooking	Ryouri	Noun	–	9.5
Driving	Unten	Noun	–	9.4
Disappointment	Shitsubou	Noun	–	9.4
Good	Yoi	Adj	–	9.4
Surprising (things)	Sapuraizu	Noun	–	9.2
Wound	Kega	Noun	–	9.2
Handy	Kiyou	Adj V	–	9.0
Feelings	Kimochi	Noun	–	8.8
Run	Hashiru	Verb	–	8.7
Exercise	Undou	Noun	–	8.7
Often	Yoku	Adv	–	8.7
Move	Hikkosu	Verb	–	8.5
Be able to take/ride	Noreru	Verb	–	8.5
Perhaps	Kamo	Particle	–	8.4
Expectation	Kitai	Noun	–	8.1
Hand fingers/dexterity	Tesaki	Noun	–	8.0
Behavior	Koudou	Noun	–	8.0
Mind	Ki	Noun	–	7.7
Go (around) with	Tsukiau	Verb	–	7.5
Lose	Makeru	Verb	–	7.5
A person’s old age	Rougo	Noun	–	7.4
Dislike	Kirai	Adj V	–	7.2
This year	Kotoshi	Adv	–	7.1
House	Ie	Noun	–	7.0
Listen	Kiku	Verb	–	7.0
Story	Hanashi	Noun	–	7.0
Build	Tateru	Verb	–	6.9
Match	Shiai	Noun	–	6.9
When	Itsu	Noun	–	6.8
Thing	Koto	Noun	–	6.8
Cheer	Genki	Adj V	–	6.6
Good (Notation variability)	Ii	Adj	–	6.5
Lose	Ushinau	Verb	–	6.3
Announcement/show	Hirou	Verb	–	6.3
Open	Hiraku	Verb	–	6.3
Fight	Kenka	Noun	–	6.3
Assessment	Kensa	Noun	–	6.3
Mother	Haha-oya	Noun	–	6.3
Dealing/coping with	Taiou	Noun	–	6.3
Lover/boy-/girlfriend	Koibito	Noun	–	6.3
Small	Chiisai	Adj	–	6.2
Student	Gakusei	Noun	–	5.9
Come	Kuru	Verb	–	5.8
Tumble	Korobu	Verb	–	5.8
Homecoming	Kaeri	Noun	–	5.8
Support/cheering	Ouen	Noun	–	5.8
Person in charge	Tantou	Noun	–	5.8
Friend	Yuujin	Noun	–	5.8
Cute	Kawaii	Adj	–	5.6

Table 8 (continued)

Morpheme (English)	Morpheme (Japanese)	Part of Speech	Chi-Squared	
			Study 1	Study 2
Relatives	Shinseki	Noun	–	5.6
Grandmother	Sobo	Noun	–	5.6
Participation	Sanka	Noun	–	5.3
Way of living	Ikikata	Noun	–	5.3
Almost	Hotondo	Adv	–	5.3
Normal	Futsuu	Adv	–	5.3
Go forward	Susumu	Verb	–	5.3
Pregnancy	Ninshin	Noun	–	5.3
“na” (Emphasis, prohibition)	Na	Particle	–	5.3
How	Dou	Adv	–	5.1
Sad	Kanashii	Adj	–	5.1
Person	Hito	Noun	–	5.1
Elementary school	Syou-gakkou	Noun	–	4.8
Occasionally	Tamani	Adv	–	4.7
Unfaithfulness	Uwaki	Noun	–	4.6
Teardrop	Namida	Noun	–	4.6
Give/gift	Okuru	Verb	–	4.6
Bouquet	Hanataba	Noun	–	4.6
Exchange	Koukan	Verb	–	4.6
Tell/convey	Tsutaeru	Verb	–	4.6
Presentation	Happyou	Noun	–	4.6
Marathon	Marason	Noun	–	4.6
Work (of art)	Sakuhin	Noun	–	4.6
Mother in law	Gibo	Noun	–	4.6
Fix	Naosu	Verb	–	4.6
Middle	Tochuu	Noun	–	4.6
Teach/tell	Oshieru	Verb	–	4.6
Report	Houkoku	Noun	–	4.6
Skiing	Suki-	Noun	–	4.6
“u” (speculation)	U	Aux V	–	4.5

Auxi V = auxiliary verb; Adj = adjective; Adv = adverb; Adj V = adjectival verb

References

- Addis, D. R., Wong, A. T., & Schacter, D. L. (2007). Remembering the past and imagining the future: Common and distinct neural substrates during event construction and elaboration. *Neuropsychologia*, *45*, 1363–1377. doi:10.1016/j.neuropsychologia.2006.10.016
- Addis, D. R., Wong, A. T., & Schacter, D. L. (2008). Age-related changes in the episodic simulation of future events. *Psychological Science*, *19*, 33–41. doi:10.1111/j.1467-9280.2008.02043.x
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.
- Brittlebank, A. D., Scott, J., Williams, J. M., & Ferrier, I. N. (1993). Autobiographical memory in depression: State or trait marker? *British Journal of Psychiatry*, *162*, 118–21. doi:10.1192/bjp.162.1.118
- Chang, W., Cheng, J., Allaire, J. J., Xie, Y., & McPherson, J. (2015). shiny: Web application framework for R (R package version 0.12.2). Retrieved from <http://CRAN.R-project.org/package=shiny>
- Conway, M. A., & Pleydell-Pearce, C. W. (2000). The construction of autobiographical memories in the self-memory system. *Psychological Review*, *107*, 261–288. doi:10.1037/0033-295X.107.2.261
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*, 273–297. doi:10.1023/A:1022627411411
- Dalgleish, T., & Werner-Seidler, A. (2014). Disruptions in autobiographical memory processing in depression and the emergence of memory

- therapeutics. *Trends in Cognitive Sciences*, 18, 596–604. doi:10.1016/j.tics.2014.06.010
- D'Argembeau, A., & Van der Linden, M. (2006). Individual differences in the phenomenology of mental time travel: The effect of vivid visual imagery and emotion regulation strategies. *Consciousness and Cognition*, 15, 342–350. doi:10.1016/j.concog.2005.09.001
- Debeer, E., Hermans, D., & Raes, F. (2009). Associations between components of rumination and autobiographical memory specificity as measured by a Minimal Instructions Autobiographical Memory Test. *Memory*, 17, 892–903. doi:10.1080/09658210903376243
- De Choudhury, M., Counts, S., & Horvitz, E. (2013). Social media as a measurement tool of depression in populations. In *WebSci '13: Proceedings of the 5th Annual ACM Web Science Conference* (pp. 47–56). New York, NY: ACM. doi:10.1145/2464464.2464480
- Duff, M. C., Hengst, J. A., Tranel, D., & Cohen, N. J. (2007). Talking across time: Using reported speech as a communicative resource in amnesia. *Aphasiology*, 21, 702–716. doi:10.1080/02687030701192265
- Duff, M. C., Hengst, J. A., Tranel, D., & Cohen, N. J. (2009). Hippocampal amnesia disrupts verbal play and creative use of language in social interaction. *Aphasiology*, 23, 926–939. doi:10.1080/02687030802533748
- Esbensen, K. M., & Geladi, P. (2010). Principles of proper validation: Use and abuse of re-sampling for validation. *Journal of Chemometrics*, 24, 168–187. doi:10.1002/cem.1310
- Forman, G. (2004). *A pitfall and solution in multi-class feature selection for text classification*. Paper presented at the 21st International Conference on Machine Learning, Banff, Canada.
- Forman, G. (2007). *Feature selection for text classification*. Retrieved from www.hpl.hp.com/techreports/2007/HPL-2007-16R1.pdf
- Gotoh, F., & Ohta, N. (2001). Affective valence of two-compound kanji words. *Tsukuba Psychological Research*, 23, 45–52.
- Griffith, J. W., Kleim, B., Sumner, J. A., & Ehlers, A. (2012). The factor structure of the Autobiographical Memory Test in recent trauma survivors. *Psychological Assessment*, 24, 640–646. doi:10.1037/a0026510
- Griffith, J. W., Sumner, J. A., Debeer, E., Raes, F., Hermans, D., Mineka, S., & Craske, M. G. (2009). An item response theory/confirmatory factor analysis of the Autobiographical Memory Test. *Memory*, 17, 609–623. doi:10.1080/09658210902939348
- Griffith, J. W., Sumner, J. A., Raes, F., Barnhofer, T., Debeer, E., & Hermans, D. (2012). Current psychometric and methodological issues in the measurement of overgeneral autobiographical memory. *Journal of Behavior Therapy and Experimental Psychiatry*, 43, S21–S31. doi:10.1016/j.jbtep.2011.05.008
- Hassanali, K. N., Liu, Y., Iglesias, A., Solorio, T., & Dollaghan, C. (2014). Automatic generation of the index of productive syntax for child language transcripts. *Behavior Research Methods*, 46, 254–262. doi:10.3758/s13428-013-0354-x
- Hastings, P., Hughes, S., Magliano, J. P., Goldman, S. R., & Lawless, K. (2012). Assessing the use of multiple sources in student essays. *Behavior Research Methods*, 44, 622–633. doi:10.3758/s13428-012-0214-0
- Heron, J., Crane, C., Gunnell, D., Lewis, G., Evans, J., & Williams, J. M. G. (2012). 40,000 memories in young teenagers: Psychometric properties of the Autobiographical Memory Test in a UK cohort study. *Memory*, 20, 300–320. doi:10.1080/09658211.2012.656846
- Higuchi, K. (2004). Quantitative analysis of textual data: Differentiation and coordination of two approaches. *Sociological Theory and Methods*, 19, 101–115.
- Higuchi, K. (2015). *KH Coder 2.x reference manual*. Retrieved from http://khc.sourceforge.net/en/manual_en_v2.pdf
- Holmes, D. I. (1985). The analysis of literary style—A review. *Journal of the Royal Statistical Society: Series A*, 148, 328–341. doi:10.2307/2981893
- Ikonomakis, M., Kotsiantis, S., & Tampakas, V. (2005). Text classification using machine learning techniques. *WSEAS Transactions on Computers*, 4, 966–974.
- Joachims, T. (2005). *Text categorization with Support Vector Machines: Learning with many relevant features*. Paper presented at the 10th European Conference on Machine Learning. doi:10.1007/BFb0026683
- Kleim, B., & Ehlers, A. (2008). Reduced autobiographical memory specificity predicts depression and posttraumatic stress disorder after recent trauma. *Journal of Consulting and Clinical Psychology*, 76, 231–242. doi:10.1037/0022-006X.76.2.231
- Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, 5, 213–236. doi:10.1002/acp.2350050305
- Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval*. Cambridge, UK: Cambridge University Press.
- Matsumoto, Y. (2000). Morphological analysis system ChaSen. *Journal of Information Processing Society of Japan*, 32, 1208–1214.
- Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., & Leisch, F. (2014). e1071: Misc functions of the Department of Statistics (e1071), TU Wien (R package version 1.6-4). Retrieved from <http://CRAN.R-project.org/package=e1071>
- Mitchell, T. (1997). *Machine learning*. New York, NY: McGraw-Hill.
- Nadkarni, P. M., Ohno-Machado, L., & Chapman, W. W. (2011). Natural language processing: An introduction. *Journal of the American Medical Informatics Association*, 18, 544–551. doi:10.1136/amiajnl-2011-000464
- Neshat-Doost, H. T., Dalgleish, T., Yule, W., Kalantari, M., Ahmadi, S. J., Dyregrov, A., & Jobson, L. (2013). Enhancing autobiographical memory specificity through cognitive training: An intervention for depression translated from basic science. *Clinical Psychological Science*, 1, 84–92. doi:10.1177/2167702612454613
- Okuda, J., Fujii, T., Ohtake, H., Tsukiura, T., Tanji, K., Suzuki, K., & Yamadori, A. (2003). Thinking of the future and past: The roles of the frontal pole and the medial temporal lobes. *NeuroImage*, 19, 1369–1380. doi:10.1016/S1053-8119(03)00179-4
- Park, L., St-Laurent, M., McAndrews, M. P., & Moscovitch, M. (2011). The immediacy of recollection: The use of the historical present in narratives of autobiographical episodes by patients with unilateral temporal lobe epilepsy. *Neuropsychologia*, 49, 1171–1176. doi:10.1016/j.neuropsychologia.2011.01.042
- Peeters, F., Wessel, I., Merckelbach, H., & Boon-Vermeeren, M. (2002). Autobiographical memory specificity and the course of major depressive disorder. *Comprehensive Psychiatry*, 43, 344–350. doi:10.1053/comp.2002.34635
- Pestian, J., Nasrallah, H., Matykiewicz, P., Bennett, A., & Leenaars, A. (2010). Suicide note classification using natural language processing: A content analysis. *Biomedical Informatics Insights*, 4, 19–28.
- Raes, F., Hermans, D., de Decker, A., Eelen, P., & Williams, J. M. G. (2003). Autobiographical memory and affect regulation: An experimental approach. *Emotion*, 3, 201–206. doi:10.1037/1528-3542.3.2.201
- Raes, F., Hermans, D., Williams, J. M. G., & Eelen, P. (2007). A sentence completion procedure as an alternative to the Autobiographical Memory Test for assessing overgeneral memory in non-clinical populations. *Memory*, 15, 495–507. doi:10.1080/09658210701390982
- Raes, F., Williams, J. M. G., & Hermans, D. (2009). Reducing cognitive vulnerability to depression: A preliminary investigation of Memory Specificity Training (MEST) in inpatients with depressive symptomatology. *Journal of Behavior Therapy and Experimental Psychiatry*, 40, 24–38. doi:10.1016/j.jbtep.2008.03.001
- Rakotomamonjy, A. (2004). Optimizing area under ROC curve with SVMs. In *Proceedings of the European Conference on Artificial Intelligence, 1st Workshop on "ROC Analysis in AI"* (pp. 469–478). Retrieved from <http://users.dsic.upv.es/~flip/ROCAI2004/papers/09-rakotomamonjyROCAI2004-revised-A4.pdf>

- Rosenbach, C., & Renneberg, B. (2015). Remembering rejection: Specificity and linguistic styles of autobiographical memories in borderline personality disorder and depression. *Journal of Behavior Therapy and Experimental Psychiatry*, *46*, 85–92. doi:10.1016/j.jbtep.2014.09.002
- Rude, S. S., Gortner, E. M., & Pennebaker, J. W. (2004). Language use of depressed and depression-vulnerable college students. *Cognition and Emotion*, *18*, 1121–1133. doi:10.1080/02699930441000030
- Sebastiani, F. (2001). Machine learning in automated text categorization. *ACM Computing Surveys*, *34*, 1–47. doi:10.1145/505282.505283
- Sing, T., Sander, O., Beerenwinkel, N., & Lengauer, T. (2005). ROCRC: visualizing classifier performance in R. *Bioinformatics*, *21*, 3940–3941. doi:10.1093/bioinformatics/bti623
- Statnikov, A., Aliferis, C. F., & Hardin, D. P. (2011). *A gentle introduction to support vector machines in biomedicine: Theory and methods*. Hackensack, NJ: World Scientific.
- Stirman, S. W., & Pennebaker, J. W. (2001). Word use in the poetry of suicidal and nonsuicidal poets. *Psychosomatic Medicine*, *63*, 517–522.
- Sumner, J. A., Griffith, J. W., & Mineka, S. (2010). Overgeneral autobiographical memory as a predictor of the course of depression: A meta-analysis. *Behaviour Research and Therapy*, *48*, 614–625. doi:10.1016/j.brat.2010.03.013
- Sung, Y. T., Chen, J. L., Cha, J. H., Tseng, H. C., Chang, T. H., & Chang, K. E. (2015). Constructing and validating readability models: The method of integrating multilevel linguistic features with machine learning. *Behavior Research Methods*, *47*, 340–354. doi:10.3758/s13428-014-0459-x
- Takano, K., Ueno, M., Mori, M., Nishiguchi, Y., & Raes, F. (2015). *Text-mining and machine learning on autobiographical memory test*. Paper presented at 10th Autobiographical Memory and Psychopathology Meeting, Leuven, Belgium.
- Takano, K., Ueno, M., Moriya, J., Mori, M., Nishiguchi, Y., & Raes, F. (2016). *The factor structure of the Autobiographical Memory Test in Japanese community samples*. Manuscript in preparation.
- Van Vreeswijk, M. F., & De Wilde, E. J. (2004). Autobiographical memory specificity, psychopathology, depressed mood and the use of the Autobiographical Memory Test: A meta-analysis. *Behaviour Research and Therapy*, *42*, 731–743. doi:10.1016/S0005-7967(03)00194-3
- Wang, Q., Hou, Y., Tang, H., & Wiprovnick, A. (2011). Travelling backwards and forwards in time: Culture and gender in the episodic specificity of past and future events. *Memory*, *19*, 103–109. doi:10.1080/09658211.2010.537279
- Williams, J. M. G., Barnhofer, T., Crane, C., Hermans, D., Raes, F., Watkins, E., & Dalgleish, T. (2007). Autobiographical memory specificity and emotional disorder. *Psychological Bulletin*, *133*, 122–148. doi:10.1037/0033-2909.133.1.122
- Williams, J. M. G., & Broadbent, K. (1986). Autobiographical memory in suicide attempters. *Journal of Abnormal Psychology*, *95*, 144–149. doi:10.1037/0021-843X.95.2.144