ORIGINAL RESEARCH

Impaired comprehension of metaphorical expressions in very mild Alzheimer's disease

Norimasa Fujimoto[1]
Hikaru Nakamura[2]
Tetsuya Tsuda[3]
Yosuke Wakutani[4]
Takeo Takao[4]

[1]Department of Rehabilitation, Kurashiki Heisei Hospital, Kurashiki, Okayama, Japan; [2]Department of Health and Welfare Science, Okayama Prefectural University, Soja, Okayama, Japan; [3]Department of Communication Sciences and Disorders, Prefectural University of Hiroshima, Mihara, Hiroshima, Japan; [4]Department of Neurology, Kurashiki Heisei Hospital, Kurashiki, Okayama, Japan

Background: Brain-damaged patients often have difficulty understanding non-literal language. However, whether patients with Alzheimer's disease (AD) have comprehension deficits of metaphorical expressions, in contrast with non-metaphorical (literal) expressions, remains unclear.

Patients and methods: The subjects were 40 AD patients; 20 had mild AD (17–23 points on the Mini-Mental State Examination [MMSE]), and 20 had very mild AD ($\geq$24 points). Twenty normal elderly controls were also enrolled as a control group. Thirty sentences that contained novel similes (Items) were prepared. For each Item, four explanatory choices, consisting of one correct response and three foils, were provided. The participants were asked to choose the written statement that best represented the Item's meaning. In addition, all the subjects completed the Token Test.

Results: The patients with mild AD had significantly lower scores than the normal controls on both the simile comprehension test and the Token Test. However, the patients with very mild AD exhibited significantly lower scores on the simile comprehension test, but not on the Token Test. The distributions of error types for the simile test differed between the mild AD group and the other groups. The mild AD patients made more errors that were "far" from the correct responses.

Conclusion: Patients with AD are more likely to have comprehension deficits of metaphorical expressions than comprehension deficits of non-metaphorical expressions. Pragmatic language dysfunction may precede formal language dysfunction during the progression of AD.

Keywords: non-literal language, simile, multiple-choice, pragmatics

Introduction

The decline in language function that occurs in patients with Alzheimer's disease (AD) is generally accepted to follow the sequences of semantics, syntax, and phonology.[1,2] In addition, linguistic communication involves not only the use of literal language but also non-literal (figurative) language, such as metaphors, similes, proverbs, idioms, and irony. Researchers examining the understanding of metaphorical expressions (metaphors and similes) have discovered deficits in patients with neurological and psychiatric diseases, such as cerebrovascular diseases involving right hemisphere lesions,[3,4] traumatic brain injury,[5,6] and schizophrenia.[7–9] Comprehension deficits of metaphorical expressions have also been examined in patients with AD, although the number of reports is small.

The first study was conducted by Winner and Gardner.[10] They examined the comprehension and explication ability of metaphors in seven AD patients. On the comprehension test, sentences that contained metaphors were verbally presented. For each sentence, the patients were asked to point to the one picture that represented the appropriate meaning of the sentence among four possible pictures. On the explication test,

Correspondence: Hikaru Nakamura
Department of Health and Welfare Science, Okayama Prefectural University, 111 Kuboki, Soja, Okayama 719-1197, Japan
Tel/fax +81 86 6 94 2181
Email hikaru@fhw.oka-pu.ac.jp

the patients were requested to explain the meaning of the metaphor within the sentence. The AD patients exhibited obviously lower scores than age-matched neurologically unimpaired controls on both tests. This study suggested the presence of metaphor comprehension deficits in AD; however, only a small number of patients were investigated, and the stage of AD was not reported.

Papagno[11] asked 39 early AD patients to perform a metaphor explication test and the Token Test,[12,13] which is a popular test for measuring comprehension of literal language. They found that only two of the 39 patients had impaired metaphor test scores and normal Token Test scores. She concluded that language comprehension impairments specific to figurative language were not an early symptom of AD. However, it should be noted that the metaphors used in both Papagno's[11] and Winner and Gardner's[10] study were conventional ones (see also Rapp and Wild[14]). It has become well known that metaphors should be divided into two types: conventional (familiar, salient) metaphors and novel (unfamiliar, non-salient) metaphors. Conventional metaphors are similar to idioms; in other words, these metaphors are frequently used as idioms, and these two types of metaphors are thought to be processed differently in the human brain.[14–17]

Maki et al,[18] on the other hand, mainly used novel metaphors and similes (see also Adachi et al[19]) to explore comprehension deficits of metaphorical expressions in 30 patients with mild AD. The patients' mean Mini-Mental State Examination (MMSE)[20] score was 21.4. In their study, written sentences that contained the metaphorical expressions were prepared. The test was a multiple-choice task. For each sentence, the patients were requested to point to the best answer that described the appropriate meaning among the five possible sentences. As a result, the patients exhibited significantly lower scores than normal elderly individuals. The patients also exhibited significantly more errors consisting of literal interpretations of the metaphorical expressions than the normal individuals. The authors claimed that literal interpretations of non-literal language were a characteristic of mild AD. Amanzio et al[21] studied relationships between conventionality of metaphors and comprehension deficits in AD. They presented both conventional and novel metaphors to 20 patients with mild AD (mean MMSE score=21.8) and asked the patients to provide a verbal explanation. They found that the patients performed significantly worse than age-matched healthy controls when tested using the novel metaphors, but not when tested using the conventional metaphors. These two studies suggest that patients with mild AD had difficulty understanding novel metaphors.

However, in both studies, the authors only compared comprehension performance for metaphorical expressions in their AD patients against that in the controls. They did not compare comprehension performance for non-metaphorical (literal) expressions between the AD and control groups. Therefore, whether comprehension deficits in AD patients are specific to metaphorical expressions remains uncertain. In addition, examining different stages of AD, especially the very early stage, may be useful for answering the above question. If a discrepancy between the comprehension of metaphorical and non-metaphorical language exists, very mild AD patients may present such a discrepancy in an obvious form. Researchers have noted that the pattern of impairment during the early stage of AD is not yet clear.[14]

The purpose of the present study was to clarify whether language comprehension deficits specific to metaphorical expressions exist in AD patients. To resolve this question, the comprehension of both metaphorical and non-metaphorical expressions was examined in patients with mild or very mild AD. Of note, the metaphorical expressions used in this study were similes, not metaphors. For example, we used the metaphorical sentence "perfume is like a bouquet," not "perfume is a bouquet." In our pilot study, some AD patients responded, "What? I cannot understand" to sentences such as "perfume is a bouquet." Thus, similes seemed to be easier to identify as non-literal expressions than metaphors.

Patients and methods
Subjects
Forty patients with probable AD, 20 with mild AD, and 20 with very mild AD participated in this study. All the subjects were outpatients attending the first author's hospital. The diagnosis of probable AD was made according to the criteria developed by the National Institute on Aging-Alzheimer's Association. Other inclusion criteria were as follows: 1) native Japanese speaker; 2) 12 or more years of education; 3) 80 years old or younger; 4) clear consciousness and stable general condition; 5) no severe psychiatric symptoms; 6) no severe visual or hearing disorders; 7) an MMSE score of 17 or more (not moderate or severe AD: Hodges and Patterson[22]); 8) a score of 6 or more on the complex sentence comprehension subtests (both auditory and visual) of the Standard Language Test of Aphasia (SLTA); and 9) consent to participate in the study. The SLTA is a comprehensive aphasia test that is commonly used in Japan. The above scores indicate that all the AD patients had equal or higher scores than those of patients with mild aphasia.[23]

Table 1 Demographic and neuropsychological data of the participating groups

	Normal controls	Very mild AD	Mild AD	P-value
Age (years)	73.6±4.7	72.3±4.0	74.5±3.9	ns[a]
Sex (male/female)	9/11	9/11	6/14	ns[b]
SLTA: Auditory word comprehension (/10)	–	8.8±1.2	9.0±1.2	ns[c]
Visual word comprehension (/10)	–	9.4±0.9	9.1±1.1	ns[c]
Picture naming (/20)	–	19.4±0.9	18.9±1.3	ns[c]
MMSE (/30)	–	25.2±1.1	20.8±1.9	<0.001[c]

Note: Values are mean±SD or numbers; [a]one-way analysis of variance; [b]chi-squared test; [c]t-test.
Abbreviations: AD, Alzheimer's disease; MMSE, Mini-Mental State Examination; ns, not significant; SLTA, Standard Language Test of Aphasia.

Mild AD and very mild AD were classified in accordance with Hodges and Patterson:[22] 17–23 points on the MMSE was classified as mild AD, and 24 points or more was classified as very mild AD.

In addition, 20 community dwelling healthy elderly subjects were also asked to participate as normal controls (NC). These subjects were recruited based on the age and sex of the AD patients. Table 1 shows the demographic characteristics of the groups. No statistically significant differences in mean age or sex distribution were seen among the groups.

Written informed consent was obtained from all the participants. The present study was approved by the ethics committee of Kurashiki Heisei hospital. The study was conducted in accordance with the principles of the Declaration of Helsinki.

Stimuli

Our simile comprehension test was comprised of 30 questions. All the similes in the test were presented in the simple form "An A is like a B" (Items). The similes were collected based on a study reported by Nakamoto and Kusumi.[24] They gathered unfamiliar metaphorical expressions from novels and apprised their originality and intelligibility using scales from 1 to 9 points in a cohort of university or college students. We selected 30 similes that had scores of 3.5 points or higher for originality and 4.0 points or higher for intelligibility. These cut-offs were applied because, in our pilot study, some similes that did not reach these points seemed to be classified as conventional ones and/or were difficult to understand even for students.

For each Item, four choices were prepared in accordance with previous studies:[25–27] the correct answer and three foils (Tenor, Vehicle, and Magical). The Tenor choice was a sentence that explained A, while the Vehicle choice was a sentence that explained B. The Magical choice was a sentence in which "An A becomes a B." For example, for the Item "perfume is like a bouquet," the Correct choice was "perfume smells good," the Tenor choice was "perfume is a fragrant liquid," the Vehicle choice was "perfume is a bundle of flowers," and the Magical choice was "perfume becomes a bouquet." Long sentences and the use of low-frequency words were avoided. Table 2 shows the examples (see also Table S1).

In each question, one Item and four choices were printed on A4 paper, in which the Item was presented at the top and the choices were listed below. The order of choices was randomized for each question. In addition, four practice items were created in the same manner. For the practice items, sentences with similes that had a relatively low originality and a high intelligibility were used.

Table 2 Examples of the simile comprehension test

Items	Choices			
	Correct	Tenor	Vehicle	Magical
Perfume is like a bouquet.	Perfume smells good.	Perfume is a fragrant liquid.	Perfume is a bundle of flowers.	Perfume becomes a bouquet.
A rumor is like a snowball.	A rumor grows rapidly.	A rumor spreads widely.	A rumor is cold and white.	A rumor becomes a snowball.
Wheat fields are like the sea.	Wheat fields are very big.	Wheat fields grow grain.	Wheat fields are full of water.	Wheat fields become the sea.
The earth is like a mother.	The earth creates life.	The earth is a large land.	The earth is a woman.	The earth becomes a mother.

Procedure

The subjects were asked to choose one written statement that best represented the Item's meaning from among the choices. For each practice exercise and question, the examiner presented the printed paper and read aloud the Item and all the choices, then asked for the subject's response. No time limit for the response was set. First, the practice exercises were performed. None of the subjects had difficulty understanding what was requested of them. Then, the test was performed. The question order was the same for all the subjects. The subjects were asked to pick one of the choices. If the subjects did not respond within about 10 seconds, the examiner read the Item and all the choices aloud a second time. After the test, all the subjects were also requested to complete the Japanese version of the Token Test[28] that was developed based on Spellacy and Spreen's[13] version. The maximum score on the Token Test is 166.

Results

Table 3 shows the percentages of correct responses on the simile comprehension test and the Token Test in each group. For the simile comprehension test, a one-way analysis of variance (ANOVA) revealed that the percentages of correct responses differed significantly among the groups (F[2, 57]=40.7; $P<0.001$). Multiple comparisons using the Tukey HSD revealed that the differences were observed among all the dyads, that is, between the NC and very mild AD groups ($P=0.01$), the NC and mild AD groups ($P<0.001$), and the very mild AD and mild AD groups ($P<0.001$). Also on the Token Test, an ANOVA revealed that the percentage of correct responses was significantly different among the groups (F[2, 57]=10.4; $P<0.001$). However, a multiple comparisons analysis revealed that significant differences were only observed between the NC and mild AD groups ($P<0.001$). The differences between the NC and very mild AD and between the very mild AD and mild AD groups were not significant ($P=0.08$ and $P=0.06$, respectively).

Table 4 shows the numbers of error responses on the simile comprehension test in each group. The chi-squared test revealed that the distributions of error responses differed significantly among the groups ($\chi^2[4]=30.0$; $P<0.001$). A residual analysis revealed that the NC and very mild AD groups exhibited significantly fewer Magical responses (both $P<0.01$), whereas the mild AD group exhibited significantly more Magical responses ($P<0.01$) and fewer Tenor responses ($P<0.01$).

Discussion

In this study, we created a simile comprehension test using novel similes and examined the performance of patients with mild or very mild AD and normal elderly controls. The results revealed that the AD patients, even the very mild AD patients, exhibited significantly lower scores than the normal elderly controls on the simile comprehension test. On the other hand, no significant differences in the Token Test scores were observed between the very mild AD patients and the normal controls. These findings suggest that the simile comprehension test might be more sensitive than the Token Test for the detection of AD.

For the Token Test, the subject was asked to point or move "tokens" (small/large and circle/rectangle pieces with white/black/red/yellow/blue colors) according to the tester's oral commands; for example, the tester asked the subject to "take the blue small circle and the yellow large rectangle" or "put the black rectangle beside the red circle." This test has been used in numerous studies on language and language disorders[29] and considered to be one of the most common tests for assessing "formal" language function, such as phonology, vocabulary, and syntax. On the other hand, metaphorical expressions require an understanding of the inferred "true" meaning, even if the phonology, vocabulary, and syntax are similar to a non-metaphorical expression. In other words, "pragmatic" language function is needed. Our results suggest that the comprehension of metaphorical expressions is more vulnerable to AD than the comprehension of non-metaphorical expressions. No significant differences in the word comprehension scores of the SLTA or the Token Test scores was seen between the very mild AD group and

Table 3 Percentages of correct responses on the simile comprehension test and the Token Test in each group

	Normal controls	Very mild AD	Mild AD	P-value
Simile comprehension test	87.5±9.1 (70.0–100.0)	76.3±10.0 (60.0–100.0)	54.8±14.9 (23.3–76.7)	<0.001[a]
Token Test	98.9±1.0 (96.4–100.0)	96.7±1.8 (92.2–100.0)	94.3±5.1 (82.0–100.0)	<0.001[b]

Note: Values are mean±SD (minimum–maximum); [a]one-way analysis of variance: significant difference between normal controls and very mild AD ($P=0.01$) and very mild AD and mild AD ($P<0.001$) with the Tukey HSD; [b]one-way analysis of variance: significant difference between normal controls and mild AD ($P<0.001$) with the Tukey HSD. **Abbreviation:** AD, Alzheimer's disease.

Table 4 Total numbers of error responses on the simile comprehension test in each group

	Normal controls	Very mild AD	Mild AD	P-value
Tenor	44 (59%)	84 (55%)	130 (44%)[b]	<0.001[a]
Vehicle	28 (37%)	57 (37%)	90 (31%)	
Magical	3 (4%)[b]	13 (8%)[b]	73 (25%)[c]	

Note: [a]chi-squared test; [b]significantly fewer than the predicted value (*P*<0.01) with the residual analysis; [c]significantly more than the predicted value (*P*<0.01) with the residual analysis.
Abbreviation: AD, Alzheimer's disease.

the mild AD group. However, the latter patients exhibited significantly lower scores on the simile comprehension test. Thus, pragmatic language dysfunction may precede formal language dysfunction in the progression of AD.

Our results also showed that the error distributions for the simile comprehension test differed among the groups. The frequency of Vehicle responses (eg, "perfume is a bundle of flowers" for "perfume is like a bouquet"), which is a literal interpretation of the simile sentence, was about the same in all the groups. However, compared with the normal controls and the very mild AD patients, the mild AD patients gave fewer Tenor responses ("perfume is a fragrant liquid") and more Magical responses ("perfume becomes a bouquet"). The Tenor response represents an acceptable and meaningful sentence. However, the Vehicle response and the Magical response are meaningless. In particular, the Magical response represents an unrealistic "fairy-tale" world[26,27] and does not have a metaphorical form, namely, "an A is likened to a B." Thus, the Tenor response may be the closest to the correct answer among the errors, while the Magical response may be the farthest from the correct answer. Therefore, the progression of AD may cause fewer Tenor and more Magical errors.

In this study, we used novel, not conventional, similes for the test. Previous studies exploring metaphor comprehension deficits in brain-damaged patients often did not explicitly describe what types of metaphors were used. Mashal et al[16] and Papagno et al[17] emphasized that conventional metaphors are coded in the mental lexicon, and only the retrieval of the meaning of the particular phrase is needed for understanding. However, novel metaphors require the creation of novel semantic connections for comprehension, requiring the retrieval of multiple word meanings in a particular phrase and the identification of common meanings. Mashal et al[15,16] performed fMRI studies and revealed that stronger (wider areas of) brain activations were observed for novel metaphors than for conventional metaphors. We think that novel, but

not conventional, metaphorical expressions should be used for language/communication tests of brain-damaged patients so as not to overlook any existing problem.

Conclusion

We concluded that the comprehension of metaphorical expressions was more vulnerable to AD than the comprehension of non-metaphorical expressions. Pragmatic language dysfunction may precede formal language dysfunction during the progression of AD.

However, the present study has some limitations. First, we did not compare the performance of simile sentence comprehension with non-simile sentence comprehension, but instead compared performance of the simile comprehension test with that of the Token Test. In other words, we compared comprehension performance for similes such as "perfume is like a bucket" with comprehension performance for following directions such as "put the black rectangle beside the red circle." Further studies comparing the comprehension of metaphorical and non-metaphorical expressions using the same vocabulary may be needed. Second, the stimuli that we used were similes, not metaphors. While these expressions are similar, they are not the same. Roncero and de Almeida[30] appraised five variables (Properties, Aptness, Familiarity, Conventionality, and Interpretive diversity) for pairs of metaphors (eg, "knowledge is power") and similes (eg, "knowledge is like power") in university students. The ratings did not differ for four of the variables; only the Properties rating was significantly lower for similes than for metaphors. The authors interpreted this finding as indicating that vehicles for similes had stronger associations with salient semantic properties than vehicles for metaphors. If true, patients with very mild AD might possibly exhibit normal understanding for metaphors. Further study examining the differences between comprehension of similes and metaphors in patients with AD is needed.

Acknowledgment

Disclosure

The authors report no conflicts of interest in this work.

References

1. Maxim J, Bryan K. *Language of the Elderly: A Clinical Perspective*. London: Whurr Publishers; 1994.

2. Emery VO. Language impairment in dementia of the Alzheimer type: a hierarchical decline? *Int J Psychiatry Med*. 2000;30(2):145–164.

3. Brownell HH, Simpson TL, Bihrle AM, Potter HH, Gardner H. Appreciation of metaphoric alternative word meanings by left and right brain-damaged patients. *Neuropsychologia*. 1990;28(4):375–383.

4. Rinaldi MC, Marangolo P, Baldassarri F. Metaphor comprehension in right brain-damaged patients with visuo-verbal and verbal material: a dissociation (re)considered. *Cortex*. 2004;40(3):479–490.

5. Towne RL, Entwisle LM. Metaphoric comprehension in adolescents with traumatic brain injury and in adolescents with language learning disability. *Lang Speech Hear Serv Sch*. 1993;24(2):100–107.

6. Yang FG, Fuller J, Khodaparast N, Krawczyk DC. Figurative language processing after traumatic brain injury in adults: a preliminary study. *Neuropsychologia*. 2010;48(7):1923–1929.

7. Champagne-Lavau M, Stip E. Pragmatic and executive dysfunction in schizophrenia. *J Neurolinguistics*. 2010;23(3):285–296.

8. Gavilán JM, García-Albea JE. Theory of mind and language comprehension in schizophrenia: poor mindreading affects figurative language comprehension beyond intelligence deficits. *J Neurolinguistics*. 2011; 24(1):54–69.

9. Mo S, Su Y, Chan RC, Liu J. Comprehension of metaphor and irony in schizophrenia during remission: the role of theory of mind and IQ. *Psychiatry Res*. 2008;157(1–3):21–29.

10. Winner E, Gardner H. The comprehension of metaphor in brain-damaged patients. *Brain*. 1977;100(4):717–729.

11. Papagno C. Comprehension of metaphors and idioms in patients with Alzheimer's disease: a longitudinal study. *Brain*. 2001;124(Pt 7): 1450–1460.

12. De Renzi E, Vignolo LA. The Token test: a sensitive test to detect receptive disturbances in aphasics. *Brain*. 1962;85(4):665–678.

13. Spellacy FJ, Spreen O. A short form of the Token test. *Cortex*. 1969; 5(4):390–397.

14. Rapp AM, Wild B. Nonliteral language in Alzheimer dementia: a review. *J Int Neuropsychol Soc*. 2011;17(2):207–218.

15. Mashal N, Faust M, Hendler T. The role of the right hemisphere in processing nonsalient metaphorical meanings: application of principal components analysis to fMRI data. *Neuropsychologia*. 2005;43(14): 2084–2100.

16. Mashal N, Faust M, Hendler T, Jung-Beeman M. An fMRI investigation of the neural correlates underlying the processing of novel metaphoric expressions. *Brain Lang*. 2007;100(2):115–126.

17. Papagno C, Lucchelli F, Muggia S, Rizzo S. Idiom comprehension in Alzheimer's disease: the role of the central executive. *Brain*. 2003; 126(Pt 11):2419–2430.

18. Maki Y, Yamaguchi T, Koeda T, Yamaguchi H. Communicative competence in Alzheimer's disease: metaphor and sarcasm comprehension. *Am J Alzheimers Dis Other Demen*. 2013;28(1):69–74.

19. Adachi T, Hirabayashi S, Shiota M, et al. [Study of situational recognition of attention deficit/hyperactivity disorders, Asperger's disorder and high functioning autism with the Metaphor and Sarcasm Scenario Test (MSST)]. *No To Hattatsu*. 2006;38(3):177–181. Japanese.

20. Folstein MF, Folstein SE, McHugh PR. "Mini-Mental State": a practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res*. 1975;12(3):189–198.

21. Amanzio M, Geminiani G, Leotta D, Cappa S. Metaphor comprehension in Alzheimer's disease: novelty matters. *Brain Lang*. 2008; 107(1):1–10.

22. Hodges JR, Patterson K. Is semantic memory consistently impaired early in the course of Alzheimer's disease? neuroanatomical and diagnostic implications. *Neuropsychologia*. 1995;33(4):441–459.

23. Japan Society for Higher Brain Dysfunction. *Standard Language Test of Aphasia Manual*. 2nd ed. Tokyo: Shinkoh Igaku Shuppansha Co, Ltd; 2003. Japanese.

24. Nakamoto K, Kusumi T. A classification of 120 Japanese metaphorical expressions on the basis of four psychological dimensions. *The Science of Reading*. 2004;48:1–10. Japanese.

25. Richards IA. *The Philosophy of Rhetoric*. New York: Oxford University Press; 1965.

26. Winner E, Rosenstiel AK, Gardner H. The development of metaphoric understanding. *Dev Psychol*. 1976;12(4):289–297.

27. Sawa T, Yoshino T. An experimental study of metaphor comprehension in children with hearing impairments. *Jap J Spec Educ*. 1994;31(4): 19–26. Japanese.

28. Hiraguchi M. *Japanese Version of the Token Test*. Kyoto: Sankyobo Co, Ltd; 2010. Japanese.

29. Lezak MD, Howieson DB, Bigler ED, Tranel D. *Neuropsychological Assessment*. 5th ed. New York: Oxford University Press; 2012.

30. Roncero C, de Almeida RG. Semantic properties, aptness, familiarity, conventionality, and interpretive diversity scores for 84 metaphors and similes. *Behav Res*. 2015;47(3):800–812.

Supplementary material

Table S1 Questions of the simile comprehension test

Items	Choices			
	Correct	**Tenor**	**Vehicle**	**Magical**
A smile is like a flower.	A smile is beautiful and lovely.	A smile is a kind of facial expression.	A smile bears fruit.	A smile becomes a flower.
A problem is like a wall.	A problem obstructs my way.	A problem is a suffering.	A problem surrounds something.	A problem becomes a wall.
A train is like a matchbox.	A train has a square shape.	A train runs with steam.	A train is small.	A train becomes a matchbox.
Perfume is like a bouquet.	Perfume smells good.	Perfume is a fragrant liquid.	Perfume is a bundle of flowers.	Perfume becomes a bouquet.
Cigarettes are like time bombs.	Cigarettes cause physical problems in the future.	Cigarettes smoke badly.	Cigarettes explode on time.	Cigarettes become time bombs.
A demonstration march is like an avalanche.	A demonstration march is a surging crowd.	A demonstration march is a citizen protest.	A demonstration march flows downwards.	A demonstration march becomes an avalanche.
A rumor is like a snowball.	A rumor grows rapidly.	A rumor spreads widely.	A rumors is cold and white.	A rumor becomes a snowball.
Fate is like a phantom killer.	Fate appears unexpectedly.	Fate exceeds one's will.	Fate can injure a person.	Fate becomes a phantom killer.
Delight is like a fountain.	Delight is a rising emotion.	Delight is a great joy.	Delight has a lot of water.	Delight becomes a fountain.
Hope is like a light.	Hope makes people's hearts bright.	Hope reflects what people want to happen.	Hope dazzles people's eyes.	Hope becomes a light.
A suspicion is like a tumor.	A suspicion can spread.	A suspicion is doubt.	A suspicion is a kind of disease.	A suspicion becomes a tumor.
Power is like drugs.	Power drives people mad.	Power dominates people.	Power is a kind of poison.	Power becomes drugs.
Labor is like a prison.	Labor restricts people's freedom.	Labor is performed by adults.	Labor is a facility to shut people in.	Labor becomes a prison.
A kitten is like a princess.	A kitten is a cute and beautiful animal.	A kitten is a frisky animal.	A kitten is an animal of high rank.	A kitten becomes a princess.
A dictionary is like a gold mine.	A dictionary is a treasure of knowledge.	A dictionary is a book in which one can look up words.	A dictionary is a mountain that produces gold.	A dictionary becomes a gold mine.
Loneliness is like a wintry blast.	Loneliness is very painful.	Loneliness is unsatisfactory.	Loneliness is a strong wind.	Loneliness becomes a wintry blast.
A shock is like electricity.	A shock is transmitted in a moment.	A shock is a kind of force.	A shock is invisible.	A shock becomes electricity.
A referee is like a balance.	A referee is unbiased.	A referee decides victory or defeat.	A referee measures weight.	A referee becomes a balance.
Wheat fields are like the sea.	Wheat fields are very big.	Wheat fields grow grain.	Wheat fields are full of water.	Wheat fields become the sea.
A man is like a wolf.	A man is dangerous and frightening.	A man can never bear a child.	A man is a canine.	A man becomes a wolf.
Roads are like blood vessels.	Road extends everywhere.	Roads are passageways.	Roads are tubes that carry blood.	Roads become blood vessels.
The earth is like a mother.	The earth creates life.	The earth is a large land.	The earth is a woman.	The earth becomes a mother.
Criticism is like a scalpel.	Criticism hurts people sharply.	Criticism points out a problem.	Criticism is used by a doctor.	Criticism becomes a scalpel.

Table S1 (*Continued*)

Items	Choices			
	Correct	**Tenor**	**Vehicle**	**Magical**
A flying fish is like a grasshopper.	A flying fish can jump.	A flying fish has a strange shape.	A flying fish eats plants.	A flying fish becomes a grasshopper.
Charm is like a magnet.	Charm attracts people's hearts.	Charm makes people crazy.	Charm attracts iron.	Charm becomes a magnet.
Fog is like a cloak.	Fog obscures objects.	Fog contains many drops of water.	Fog is a type of clothing without sleeves.	Fog becomes a cloak.
A sunset is like a copper coin.	A sunset shines red.	A sunset glows in the evening.	A sunset is money.	A sunset becomes a copper coin.
A theory is like a building.	A theory is a carefully constructed matter.	A theory is a carefully considered matter.	A theory is a type of man-made object.	A theory becomes a building.
Shooting stars are like raindrops.	Shooting stars fall downwards.	Shooting stars burn and disappear.	Shooting stars are composed of water.	Shooting stars become raindrops.
An old man is like a piece of furniture.	An old man does not move.	An old man is a senior citizen.	An old man is in a house.	An old man becomes a piece of furniture.