Listeners’ identification and evaluation of Korean idiomatic utterances produced by persons with left- or right-hemisphere damage

Seung-yun Yang, Diana Sidtis & Seung Nam Yang

To cite this article: Seung-yun Yang, Diana Sidtis & Seung Nam Yang (2017) Listeners’ identification and evaluation of Korean idiomatic utterances produced by persons with left- or right-hemisphere damage, Clinical Linguistics & Phonetics, 31:2, 155-173, DOI: 10.1080/02699206.2016.1186223

To link to this article: http://dx.doi.org/10.1080/02699206.2016.1186223

Published online: 24 Jun 2016.
Listeners' identification and evaluation of Korean idiomatic utterances produced by persons with left- or right-hemisphere damage

Seung-yun Yang\textsuperscript{a,b}, Diana Sidtis\textsuperscript{b,c}, and Seung Nam Yang\textsuperscript{d}

\textsuperscript{a}Graduate Program in Speech Language Pathology, Touro College, Brooklyn, NY, USA; \textsuperscript{b}Brain and Behavior Laboratory, Geriatrics Division, Nathan Kline Institute for Psychiatric Research, Orangeburg, NY, USA; \textsuperscript{c}Department of Communicative Sciences \& Disorders, New York University, New York, NY, USA; \textsuperscript{d}Department of Physical Medicine and Rehabilitation, Korea University College of Medicine, Korea University Guro Hospital, Seoul, South Korea

\begin{abstract}
This study investigates the effects of left- (LHD) or right-hemisphere damage (RHD) on the production of matched idiomatic or literal expressions by examining healthy listeners’ abilities to identify, evaluate and perceptually characterize the utterances. Native speakers of Korean with LHD or RHD and healthy controls (HCs) produced six ditropically ambiguous (idiomatic or literal) sentences in an elicitation and a repetition task. Healthy listeners identified the sentence types and indicated how well each utterance represented the intended meaning. Perceptual ratings of voice quality were performed by expert listeners. The results indicate a negative effect of RHD on listeners’ identification and goodness ratings of utterance type. Repetition yielded better speech exemplars than elicitation. Sentence type was associated with selected voice qualities. These findings support previous reports of prosodic information serving to signal idiomatic versus literal meanings as well as a right-hemisphere involvement in formulaic language and the dual process model of language.
\end{abstract}

\section*{Introduction}

\textbf{Background: Idiom studies}

Idioms hold special interest in the linguistic sciences. Despite decades of inquiry, it remains rather mysterious to consider that by using idioms, speakers are able to convey ideas and emotions indirectly, because idioms do not rely on standard lexical meanings to make their statement. \textit{She has him eating out of her hand} does not reference eating or hands, but indexes an asymmetrical sociopolitical relationship between two persons of different genders, and carries connotations of domination, submission, compliance, with possible nuances of arrogance and victimization. That idioms use words differently than word usage in literal or propositional utterances is well known to linguists and laypersons alike.
Studies of idioms reflect a venerable expanse of activity across Europe and North America since the early part of the twentieth century. Along with their companion category, proverbs, idioms have stimulated numerous questions about the processes by which they communicate meanings beyond the usual lexical semantics of their constituent words. Early studies dealt with identifying, defining and characterizing idioms (Hockett, 1958; Makkai, 1972; Roberts, 1944; Smith, 1925). With the advent of psycholinguistics later in the century, language scientists examined storage and processing, addressing questions about structure and constituency (e.g. Bobrow & Bell, 1973; Gibbs & Gonzales, 1985; Peterson, Burgess, Dell, & Eberhard, 2001; Sprenger, Levelt & Kempen, 2006; Swinney & Cutler, 1979; Titone & Connine, 1999). Several theories emerged to usefully address problems about the potential compositionality or noncompositionality of idioms (Titone & Connine, 1999), and although controversies remain about their ‘holistic’ structure (Siyanova-Chanturia, 2015; see Van Lancker Sidtis, 2010), it is generally agreed that at a demonstrable level of mental organization, idioms are known and processed in a unitary fashion (Rammell, Pisoni, & Van Lancker Sidtis, 2016; Sprenger et al., 2006; Swinney & Cutler, 1979).

The current study is anchored in another, essential assumption about idioms—not often clearly stated in the airing of idiom linguistics—that people know them (Kuiper, Van Egmond, Kempen, & Sprenger, 2007). This intuitively obvious fact has been empirically verified in formal studies (Hallin & Van Lancker Sidtis, 2015; Sprenger, 2003; Van Lancker Sidtis, Cameron, Bridges, & Sidtis, 2015; Van Lancker Sidtis, Kougentakis, Cameron, Falconer, & Sidtis, 2012; Van Lancker Sidtis & Rallon, 2004; Vaynshteyn, 2015). We assert further that formulaic expressions are known by native speakers of a language in terms of their canonical shape and their attendant semantic and contextual characteristics: Stereotyped form with unique phonetic/prosodic details, conventional, non-literal meanings carrying affective nuances, and a unique relationship to context. These properties form a distinctive set in comparison to the properties inhering in newly created, propositional utterances. Native speakers ‘know’ the expression, *She has him eating out her hand*, in a way that has no application to a newly generated utterance, *He takes his pets in the car*.

**Auditory-acoustic cues associated with idiomatic expressions**

As mentioned above, idiomatic expressions are characterized by stereotyped form including specific words in a fixed order. Such expressions often carry signature prosodic properties, such as canonical intonation contour, specific sentence accents, word lengths and distinctive voice quality (Lin, 2010; Van Lancker Sidtis, 2008). It was noted some time ago that the articulatory and resultant perceptual characteristics of these two types of expressions differ from each other (Lieberman, 1963). A previous study (Van Lancker, Canter, & Terbeek, 1981) also reported that idiomatic utterances differ from literal utterances with respect to measurable speech characteristics and perceived details of voice quality. An analysis of acoustic cues of idiomatic and literal sentences in the earlier studies revealed that the two types of utterances were significantly different in durational and fundamental frequency measures (Yang, Ahn, & Van Lancker Sidtis, 2015) as well as voice quality (Van Lancker et al., 1981), such as creaky voice or murmur, compared with literal counterparts.
Identifying the intended meanings of ambiguous idiomatic expressions based solely upon auditory-acoustic characteristics has been reported as reflecting a native speaker’s competence. Healthy listeners correctly identified sentences as either literal or idiomatic at a level well above chance from the acoustic signal alone, in the absence of any situationally or linguistically disambiguating context (Van Lancker & Canter, 1981); these results were found also for Korean sentences (Yang et al., 2015). For the US English stimuli, it was further shown that native listeners performed significantly better than non-native listeners regardless of the amount of language exposure and training non-native listeners received, suggesting that identifying these specific prosodic contrasts is part of a native speaker’s competence (Van Lancker Sidtis, 2003).

*Cerebral processing of idioms*

A central question in this study addresses brain processing of formulaic expressions. For comprehension, variable results have appeared. A right-hemisphere involvement in the comprehension of non-literal expressions was supported by a number of earlier investigations on pragmatic and inferential deficits in individuals with right-hemisphere damage (RHD) (Joanette, Goulet, & Hannequin, 1990; Molloy, Brownell, & Gardner, 1990). Individuals with RHD exhibited impaired performance on tasks that required the understanding of idioms (Van Lancker & Kempler, 1987; Van Lancker Sidtis, 2006; Winner & Gardner, 1977). However, several recent investigations implicate a left-hemisphere involvement in idiom comprehension (Papagno, Curti, Rizzo, Crippa, & Colombo, 2006; Papagno, Oliveri, & Romero, 2002; Papagno, Tabossi, Columbo, & Zampetti, 2004). Some of these discrepancies might be explained by the kinds of tasks included in the study designs. For comprehending the meanings of idioms, a clear picture of hemispheric specialization has not emerged, and it may be that each hemisphere plays its unique role when comprehension is examined.

For language production, evidence elucidating the brain structures underlying formulaic language is somewhat clearer. It has been well known for 150 years that persons with left-hemisphere damage (LHD) retain the ability to verbally produce a great range of formulaic expressions, which are uttered with normal articulation and prosody (Wepman, Bock, Jones, & Van Pelt, 1956). Starting with Hughlings Jackson (1874), the neural structure deemed responsible for this preserved speech performance has been the right hemisphere. A formal study of the preservation of automatic speech in aphasia by Lum and Ellis (1994) was partly confirmatory of clinical observations but yielded equivocal results (for review see Van Lancker Sidtis, 2010).

More recently, long-standing clinical observations have been augmented by empirical measures, which reveal a significantly greater than normal incidence of formulaic expressions in monologues spoken by persons with left-brain damage and aphasia, while a significant paucity was associated with speech from persons with unilateral RHD due to stroke (Sidtis, Canterucci, & Katsnelson, 2009; Van Lancker Sidtis & Postman, 2006; Yang & Van Lancker Sidtis, 2015). Converging evidence from these numerous clinical observations and experimental studies leads to the conclusion that the right hemisphere makes a major contribution to the presence of formulaic language in spontaneous speech. These results contribute to a ‘dual process’ model, whereby formulaic and novel languages are modulated by differing cerebral systems. The dual
process model has been adopted as explanatory, in various forms, by several researchers in this field of study (Erman & Warren, 2000; Foster, 2001; Heine, Kuteva, & Kaltenböck, 2014; Siyanova-Chanturia & Martinez, 2015; Van Lancker Sidtis, Choi, Alken, & Sidtis, 2015; Wray and Perkins, 2000).

The current study serves our interest in cerebral processing of idioms by testing native speakers’ abilities to produce literal and idiomatic meanings in ditropically ambiguous sentences, utilizing two experimental groups: persons with LHD and those with RHD, compared to a healthy control (HC) group. The research question is focused on the effects of hemispheric damage on abilities to produce utterances with idiomatic meanings, as contrasted with matched utterance-types produced with intended literal meanings. The effect of brain damage on the production of ditropic sentences was examined perceptually through listeners’ performance, namely listeners’ abilities to identify whether the utterances produced by the study participants and the HC speakers were idiomatic or literal in intended meaning, and to evaluate the success of each production. This paper presents the results of normal listeners’ abilities to recognize, evaluate and determine voice quality on utterance-types produced by persons in these three study groups.

**Task effects on speech characteristics**

Speech characteristics differ with register, context and mode of speaking, whether spontaneous, repeated or read aloud (Andrews, Howie, Dozsa, & Guitar, 1982; Ferguson & Kewley-Port, 2007; Hébert, Racette, Gagnon, & Peretz, 2003; Ludlow & Loucks, 2003; Van Lancker Sidtis, Pachana, Cummings, & Sidtis, 2006). Task can differentially affect speech performance in patients with different types of aphasia (Goodglass & Kaplan, 1972). Measures of voice, fundamental frequency and fluency as well as of intelligibility have been reported to vary with spontaneous speech as compared with repetition in disordered speech (Kempler & Van Lancker, 2002; Van Lancker Sidtis, Rogers, Godier, Tagliati, & Sidtis, 2010; Van Lancker Sidtis, Hanson, Jackson, Lanto, Kempler, Metter, 2005). These studies indicate that performing studies on samples derived from only one kind of task, e.g. repetition, may not yield results that can be generalized to typical language use. To capture the effects of task on other independent variables of this study, speech samples were obtained in both spontaneous (elicited) and repeated modes. This design enabled a direct comparison of these two tasks utilizing the same sentences. Elicited speech, which is defined in our study as free production of sentences given contextually biased linguistic contexts, is meant to simulate spontaneous speech as closely as possible. It requires the speaker to generate an action plan and control the action using internal models. Repeated speech, however, is the repetition of the same sentences presented and modelled by the examiner. Providing an external model reduces the burden of the initiation of speech planning and control. Thus, this study examined speech production for non-literal and literal contrasting expressions using matched utterances obtained in two speech tasks in adults with LHD- or RHD, by measuring listeners’ performance in identifying, evaluating and perceptually rating the utterances.
Goals of study

While some research has been performed on the production of non-literal sentences in healthy speakers, little is known about the production of idiomatic expressions by brain-damaged individuals. This study investigated whether literally produced compared with idiomatically produced utterances carry appropriately distinguishing prosodic features in selected brain-damaged individuals. The question was whether distinctive neural circuitry is involved in the production of different prosodic cues for different types of utterances. Literal and idiomatic versions of the matched sentence pairs carry different auditory/acoustic cues. Our question was whether and in what ways hemispheric damage affected abilities to produce these differentiating cues in the spoken utterance. Our measures were listeners’ abilities to identify and evaluate these utterances.

Three performance measures were obtained from listeners. First, healthy listeners identified the category, idiomatic vs. literal, of the ditropic sentences produced by the brain-damaged participants. The aim of the identification procedure was to assess whether Korean native listeners are able to identify the utterance contrasts between idiomatic and their paired literal meanings of ambiguous sentences produced by two groups of brain-damaged individuals (in comparison to comparable stimuli elicited from healthy speakers) from the auditory-acoustic signal alone without any situational-linguistic context and presented in randomized order. Second, healthy listeners rated each stimulus in terms of its success in communicating a literal or idiomatic meaning (goodness rating). The goodness rating procedure probed how well the auditory/acoustic characteristics associated with the two different types of utterances were assessed by native listeners to represent the two different meaning types. (Correlational analyses between the results of identification and goodness ratings were conducted.) In this procedure, listeners were informed of the intended utterance category (literal or idiomatic) before hearing the utterance. The third source of perceptual data was derived from voice quality ratings by expert listeners, using scales selected to reflect cues that might differentiate literal from idiomatic categories in the stimuli. The aim of vocal quality ratings by expert listeners was to systematically investigate what voice qualities optimally categorize utterances into different types of sentences (idiomatic or literal), and to identify voice qualities which might be affected by the hemispheric damage. Perceptual ratings were performed by expert listeners on all speech samples presented randomly.

Method

Stimulus acquisition

Speakers
Speaker-participants were seven individuals with LHD, seven individuals with RHD, and seven age- and education-matched non-brain-damaged HCs. They were all right-handed native speakers of Korean, born and educated in Korea. No history of speech/language disorders or any major medical, neurological or psychiatric conditions was reported.

Individuals with brain damage. The mean ages of individuals with LHD and RHD were 62.71 and 64 with a mean education of 15.14 and 14.57. Participants with brain damage had all suffered a single unilateral lesion due to cerebrovascular accident (CVA) and
ranged in time post-onset of stroke from eight months to five years and six months with a mean of two years and seven months (LHD) and one year and nine months (RHD). Confirmation of lesion site for each participant was obtained from CT scan or MRI records.

Language abilities were determined by the Korean version of the Western Aphasia Battery (K-WAB) (Kim & Na, 2001). The K-WAB was also administered to participants with RHD to confirm that they did not exhibit any language impairment. The severity of aphasia, as determined by the Aphasia Quotient (AQ) of the K-WAB, was mild to moderate for participants with LHD and participants with RHD did not show any impairment in language. Participants with RHD underwent the Korean version of the Mini Inventory of Right Brain Injury (K-MIRBI) to detect the presence of neurocognitive deficits associated with RHD. The K-MIRBI was not administered to participants with LHD since the results can be invalid due to their speech and language difficulties. All participants with brain damage met the inclusion criteria by obtaining 50 or above AQ on the K-WAB. None of the brain-damaged participants exhibited any significant motor speech disorders affecting the speech musculature based on the evaluation by speech-language pathologists. Detailed information regarding demographic and clinical attributes of all participants is summarized in Table 1.

Non-brain-damaged HC. The mean age of the HC participants was 64.5 with a mean education of 14.8. The HC participants were administered the K-WAB to confirm that they did not exhibit any language impairment. Demographic information and the test score are presented in Table 2.

### Table 1. Information about participants with unilateral cerebral lesions.

<table>
<thead>
<tr>
<th>LHD/ RHD</th>
<th>Age (year)</th>
<th>Sex</th>
<th>Education (year)</th>
<th>Time post onset (years; months)</th>
<th>Lesion site</th>
<th>WAB (AQ)</th>
<th>MIRBI</th>
</tr>
</thead>
<tbody>
<tr>
<td>LHD</td>
<td>1</td>
<td>56</td>
<td>Female</td>
<td>16</td>
<td>0;11</td>
<td>Frontal</td>
<td>70.6</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>61</td>
<td>Male</td>
<td>12</td>
<td>1;1</td>
<td>Temporo-parietal</td>
<td>87.2</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>67</td>
<td>Male</td>
<td>16</td>
<td>5;6</td>
<td>Fronto-parietal</td>
<td>63.2</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>63</td>
<td>Female</td>
<td>12</td>
<td>1;9</td>
<td>Fronto-parietal</td>
<td>90.2</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>59</td>
<td>Female</td>
<td>16</td>
<td>4;6</td>
<td>Temporal-parietal</td>
<td>78.3</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>68</td>
<td>Male</td>
<td>16</td>
<td>3;0</td>
<td>Parietal</td>
<td>82.4</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>65</td>
<td>Male</td>
<td>18</td>
<td>0;8</td>
<td>Frontal</td>
<td>64.8</td>
</tr>
<tr>
<td>Average</td>
<td>62.71</td>
<td>15.14</td>
<td></td>
<td>2.7</td>
<td></td>
<td>76.67</td>
<td></td>
</tr>
<tr>
<td>RHD</td>
<td>1</td>
<td>63</td>
<td>Female</td>
<td>16</td>
<td>1;4</td>
<td>Temporal</td>
<td>98</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>57</td>
<td>Female</td>
<td>18</td>
<td>2;5</td>
<td>Frontal</td>
<td>99.0</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>72</td>
<td>Male</td>
<td>12</td>
<td>1;7</td>
<td>Temporal-parietal</td>
<td>94.5</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>67</td>
<td>Male</td>
<td>12</td>
<td>3;2</td>
<td>Frontal-temporal-parietal</td>
<td>96.8</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>64</td>
<td>Male</td>
<td>12</td>
<td>3;2</td>
<td>Temporal-parietal</td>
<td>100.0</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>59</td>
<td>Male</td>
<td>16</td>
<td>2;2</td>
<td>Frontal</td>
<td>95.4</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>66</td>
<td>Male</td>
<td>16</td>
<td>0;9</td>
<td>Fronto-parietal</td>
<td>98.5</td>
</tr>
<tr>
<td>Average</td>
<td>64</td>
<td>14.57</td>
<td></td>
<td>1;9</td>
<td></td>
<td>97.46/100</td>
<td>26.86/43</td>
</tr>
</tbody>
</table>

### Table 2. Information about healthy controls (WAB (AQ) = Western Aphasia Battery (Aphasia Quotient)).

<table>
<thead>
<tr>
<th>HC</th>
<th>Age (year)</th>
<th>Sex</th>
<th>Education (year)</th>
<th>WAB (AQ)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Range</td>
<td>56–71</td>
<td>3 females/4 males</td>
<td>12–18</td>
<td>98–100</td>
</tr>
</tbody>
</table>
Stimuli

Six Korean ditropic sentence pairs were utilized. They were obtained from a rating study with a cohort matched in age to the participants. Stimuli were initially selected from 20 idiomatic expressions based on familiarity, meaningfulness and literal plausibility ratings (Libben & Titone, 2008). Twenty native speakers of Korean then rated 20 corresponding Korean idiomatic phrases according to familiarity, meaningfulness and literal plausibility. Six ditropic sentences that obtained the highest ratings were utilized as stimuli. All study sentences had a subject+verb+object structure. Appendix 1 includes the detailed information about stimuli. Formal and informal morphemes were included in order to balance the syllable length in the stimulus pairs. Participants in the stimulus acquisition portion of the study were encouraged to produce the entire utterance. The idiomatic meanings of the selected stimuli were all nuanced, as is usually the case in idioms, towards a slightly negative valence. Examples in English are He’s at the end of his rope; The shoe is about to drop; hold your horses. This characteristic, having attitudinal connotations, is typical of idiomatic meanings, and contributes to their recognizability and distinguishability from any literal counterpart meaning.

Procedures

The stimuli produced by participants (RHD, LHD and HC) consisted of two different tasks, the elicitation task and the repetition task. The recordings were conducted in a quiet environment using a head-worn microphone maintaining a constant mouth-to-microphone distance of 2 inches. Microphone position and the recording gain settings were the same across all participants.

All participants were first given a training session in which it was explained that some sentences can have two different meanings, idiomatic or literal, depending on the situational-linguistic context. It has been our experience that the distinction between idiomatic and literal meanings, when examples are given, is easily understood by native speakers of a language. Sample utterances, not included in the actual elicitation tasks, were provided in the training session. During the training session, participants reviewed the list of the target sentences and became familiar with the sentences and the procedures. Each participant was encouraged to carefully examine the list of the target sentences.

During the elicitation tasks, linguistic-situational context (Appendix 1) was provided by the experimenter to elicit participants’ productions of target utterances. Situational-linguistic context was given in the form of possible scenarios both verbally and in written form. For the ambiguous sentence 그 사람이 입을 다물었어 (He closed his mouth), one linguistic context (During the police investigation, he did not say anything) referenced the idiomatic interpretation (He kept silent). The other linguistic context (After the oral surgery, he could not open his mouth) referenced the literal interpretation (He had his mouth closed) (Linguistic contexts for each target utterance utilized during the elicitation tasks are provided in Appendix 1). Participants were provided with cards written with the target sentences, their intended interpretations labelled as either ‘idiomatic’ or ‘literal’, and the respective linguistic context for the idiomatic or literal interpretation. Before participants produced the target utterances, the cards containing the written target sentences and their intended interpretations were taken away, in order to encourage participants to produce the sentences in a natural manner. Participants produced the target sentence, first, with the appropriate idiomatic interpretation and with the literal interpretation.
Participants were allowed to revisit the list of target utterances whenever they needed to. In testing, participants did not show any signs of confusion about the nature of the task. Twelve utterances were elicited (six utterances with the idiomatic interpretation and six utterances with the literal interpretation) per speaker (LHD, RHD or HC), yielding a total of 252 utterances.

The same sentences utilized for the elicitation task were used in the repetition task. Each utterance corresponding to the intended meanings (literal or idiomatic) was produced and pre-recorded by a native speaker trained to produce the utterances in a natural manner. For the repetition task, all participants were asked to repeat the sentences from the recording as closely as possible, resulting in 252 repeated utterances for each language group. Participants were informed of the intended meaning of each utterance before they repeated it.

The two stimulus acquisition procedures, elicitation and repetition, were given in that order to rule out the possible effect of pre-recorded modelled utterances on participants’ production of elicited utterances.

Identification and goodness ratings

Listeners

The data were collected from 40 native listeners of Korean with ages between 25 and 40, born and raised in South Korea. They ranged in education from 9 to 18 years. They had no hearing deficits and history of speech/language disorders and major medical, neurological and psychological conditions.

The utterances recorded by brain-damaged individuals and HCs during the stimuli acquisition process were utilized as stimuli, yielding a total of 504 stimuli (252 stimuli obtained from the elicitation task and 252 stimuli obtained from the repetition). Two listening sets were prepared, each containing only one speech task version of each utterance (i.e. elicited or repeated) for each conveyed meaning. The second listening set contained the inverse task version of each stimulus. This was so that all items were heard in both task versions, but no one listener heard both task versions (elicited and repeated) of any single stimulus. This was done to avoid practice or order effects.

Listening procedures

The experiment was designed and implemented using PRAAT (V. 4.6. 34) (Boersma & Weenink, 2007). Stimuli were presented in a quiet room over circumaural headphones. Participants were divided into four groups and randomly assigned to one of two listening sets and performed either the identification or the goodness rating tasks. Listeners were instructed to take as much time as necessary. Listening sets were given in blocks, first identification, then goodness ratings. In the identification and the goodness rating tasks, elicitation and repetition exemplars were randomized, but blocked into two mirrored protocols, so that no listener heard the same stimulus in the two task versions.

Identification. The single sentence identification task was utilized to evaluate listeners’ ability to identify utterances that are either idiomatic or literal without benefit of any contextual information. Listeners were requested to determine which meaning was intended in a two-alternative forced choice by listening to each utterance. For each
utterance, participants were given two choices (buttons) on the screen, idiomatic and literal, and instructed to click one of the buttons after hearing each utterance. Responses were scored as correct or incorrect for each utterance. Two practice items, not included in the identification procedure, were given for each sentence type in the beginning of the listening session.

**Goodness ratings.** In the second experimental procedure, goodness rating, healthy listeners were instructed to rate each item individually on a scale of 1 to 5, with 5 being a good example (of an idiomatic or a literal utterance) and 1 being a very poor example on presentation of the stimulus item. Listeners were informed, before presentation of the stimulus, about whether the given utterance was spoken with an intended idiomatic or literal meaning.

**Vocal quality: Perceptual ratings**

**Raters**
Four speech-language pathologists with training in phonetics participated as raters of voice quality in this study. The raters were native speakers of Korean, and they have had extensive experience in rating both pathological and non-pathological voices.

**Vocal quality scales**
The perceptual rating of voice quality utilized six voice quality features based on Laver (1980) and previous studies regarding vocal qualities of non-literal utterances (Bryant & Fox Tree, 2005; Paulmann, Pell, & Kotz, 2008). Certain of these voice qualities (breathy, whispery, creaky, harsh, pharyngealized, nasal/non-nasal) have been recognized to be associated with non-literal utterances such as sarcasm (Van Lancker et al., 1981). For each of the six voice quality characteristics, a 5-point equal-appearing interval scale was utilized. The scale ranged from 1 to 5, with 1 referring to the absence of the feature in question.

**Rating procedures**
Before the perceptual rating, a training session was given to all raters to familiarize them with the procedure. The training session involved reviewing definitions of the voice quality features that were used in the perceptual rating procedure. Samples for the training trials were obtained separately from healthy speakers and were not included in the actual perceptual rating. During the training trials, the listeners were encouraged to discuss with the experimenter regarding the voice quality features and rating procedure. The stimuli were presented to each rater individually in a quiet room over closed headphones. The raters were asked to perceptually judge each speech stimulus for the voice quality features on the response sheet provided. Speech stimuli were played twice, but were played again if the raters requested. Stimuli were randomized for presentation and the raters were blind to the sentence type and the status of the speakers who provided the speech samples. They were instructed to take as much time as necessary.

**Statistical analyses**
Analyses of variance (ANOVA)s in the form of $3 \times 2 \times 2$, with three groups (RHD, LHD, HC), two sentence types (literal, idiomatic) and two tasks (elicitation, repetition) were conducted to
analyse the results from the identification and goodness rating procedures. Independent and paired sample t-tests were conducted as post hoc analyses to further analyse between-group and within-group comparisons and to compare between tasks. Correlation coefficients were calculated to examine the relationship between the identification task results and goodness ratings of each utterance. Paired sample t-tests were conducted to compare between idiomatic and literal utterances with respect to each vocal quality feature. In the present study, an alpha level of 0.05 was adopted to determine statistical significance.

**Results**

**Identification**

Statistical comparisons were performed with task (elicited or repeated), sentence type (idiomatic or literal) and group (LHD, RHD, HC) as factors. A repeated measures ANOVA revealed a significant main effect of type \( F(1, 123) = 15.668; p < 0.001 \), group \( F(2, 123) = 152.900 \) and task \( F(1, 123) = 125.630; p < 0.001 \). Interactions of type × group \( F(2, 123) = 6.638; p = 0.002 \) and of task × group \( F(2,123) = 25.475; p < 0.001 \) were revealed.

For the sentence type comparison (Figure 1), literal utterances (M = 7.08, SD = 0.12) were correctly identified better than idiomatic utterances (M = 6.44, SD = 0.11), which was confirmed by a repeated measures ANOVA \( F(1, 123) = 15.668; p < 0.001 \). Comparing between groups, listeners performed better in identifying the intended meanings of utterances produced by HC (M = 8.41, SD = 0.14) compared to ones produced by brain-damaged participants. Utterances produced by the LHD group (M = 6.93, SD = 0.14) were identified better than ones by the RHD group (M = 4.93, SD = 0.14). These differences were revealed by a repeated measures ANOVA \( F(2, 124) = 152.900; p < 0.001 \).

For the comparison between speech tasks (Figure 2), elicited and repeated, listeners identified the intended meanings of ditropic sentences produced by participants with less accuracy for elicited utterances (M = 6.17, SD = 0.11) compared to repeated counterparts (M = 7.34, SD = 0.08). This difference was significant as revealed by a repeated measures ANOVA \( F(1, 123) = 125.630; p < 0.001 \).

![Figure 1](image-url)  
*Figure 1.* Means and standard errors for correct responses for elicited utterances produced by the three subject groups on the identification task. Literal utterances are represented by filled bars. Idiomatic utterances are represented by open bars. Asterisks (*) indicate statistical significance.
Paired-samples $t$-tests were performed to examine performance by each brain-damaged group on the two sentence types (literal or idiomatic) (Figure 1), revealing that listeners showed sharply decreased performance in identifying intended meanings of ditropic sentences produced by speakers with RHD, especially when presented with idiomatic-intended as compared with literal-intended utterances [$t(41) = -4.228, p < .001$]. There was no significant difference in healthy listeners’ identification of idiomatic versus literal utterances obtained from persons with LHD and HC.

Performance by brain-damaged participants on the two tasks (Figure 2) was further examined. Post hoc tests revealed that repeated utterances produced by either brain-damaged group were associated with better listeners’ performance compared to the elicitation tasks ([$t(83) = -6.02; p < 0.001$] for the LHD group; [$t(83) = -10.64; p < 0.001$] for the RHD group), which likely reflects the fact that distinctive acoustic characteristics were provided to speakers by the modelled sentences. Thus, although showing a deficit in planning and producing the utterances in the elicitation mode, the study participants were able to imitate a model. Tasks did not affect the listeners’ performances for the utterances produced by HC.

**Goodness ratings**

A repeated measures ANOVA revealed a significant main effect of type [$F(1, 123) = 11.964; p = 0.001$], group [$F(2, 123) = 218.990; p < 0.001$] and task [$F(1, 123) = 253.565; p < 0.001$]. The procedure further revealed interactions of type × group [$F(2,123) = 6.096; p = 0.003$] and of task × group [$F(2, 123) = 65.824; p < 0.001$] (Figure 3).

With respect to sentence type, literal utterances (M = 3.41, SD = 0.04) had higher ratings compared to idiomatic utterances (M = 3.22, SD = 0.04), which was also confirmed by a repeated measures ANOVA [$F(1, 123) = 11.964; p = 0.001$]. Higher ratings were given to utterances produced by HC (M = 3.97, SD = 0.05) compared to ones produced by LHD (M = 3.33, SD = 0.05) and RHD (M = 2.64, SD = 0.05). This difference was revealed by a repeated measures ANOVA [$F(2, 123) = 218.990; p < 0.001$]. Comparing between speech tasks (elicited and repeated), higher ratings were given to repeated utterances (M = 3.51,
SD = 0.02) compared to elicited utterances (M = 3.11, SD = 0.03), which was significantly different as revealed by a repeated measures ANOVA \[F(1, 123) = 253.565; \ p < 0.001\].

Post hoc tests for performance by different groups of participants on two sentence types (literal or idiomatic) revealed that for the utterances produced by RHD, significantly higher ratings were given on literal utterances compared to idiomatic utterances \[t(41) = -4.51; \ p < 0.001\]. Such differences between two sentence types were not seen in the utterances produced by LHD and HC.

Paired-samples \(t\)-tests were used to examine performance by different groups of participants on the two tasks. Repeated utterances produced by brain-damaged groups (LHD and RHD) were associated with higher ratings compared to elicited utterances ((\(t(83) = -13.39, \ p < 0.001\)) for RHD and \(t(83) = -7.331, \ p < 0.001\) for LHD), further reflecting the differential effects of speech tasks. However, tasks did not affect listeners’ ratings on utterances produced by HC.

**Relationship between identification and goodness ratings**

The relationship between percentage of correct responses from identification and goodness ratings on each utterance was examined. A significant correlation between ratings on utterances and correct responses was found for both idiomatic and literal utterances across groups during the elicitation tasks (for idiomatic utterances by LHD \(r(42) = 0.81; \ p < 0.001\); for literal utterances by LHD \(r(42) = 0.9; \ p < 0.001\); for idiomatic utterances by RHD \(r(42) = 0.8; \ p < 0.001\); for literal utterances by RHD \(r(42) = 0.92; \ p < 0.001\); for idiomatic utterances by HC \(r(42) = 0.78; \ p < 0.001\); for literal utterances by HC \(r(42) = 0.87; \ p < 0.001\)). These results revealed that higher goodness ratings on elicited utterances were strongly associated with better identification of the intended meaning of the utterances. Correlational analysis was not performed on repeated utterances due to the high accuracies in all sentence types and the presumptive role of modelling the utterances for the study participants.
Ratings of vocal quality

Of the six parameters presented to expert listeners for ratings (breathy, whispery, creaky, harsh, pharyngealized, nasal/non-nasal), only one emerged as significant in all groups of participants for both elicitation and repetition tasks: the nasal/non-nasal dimension. Elicited idiomatic utterances were judged to be nasal compared to literal utterances for RHD \( (t(125) = 9.3; p < 0.001) \); LHD \( (t(125) = 14.65; p < 0.001) \) and HC \( (t(125) = 4.28; p < 0.001) \). Similarly, for repeated utterances, the nasal/non-nasal dimension significantly differentiated the two sentence types: RHD \( (t(125) = 13.62; p < 0.001) \); LHD \( (t(125) = 18.02; p < 0.001) \); and HC \( (t(125) = 18.95; p < 0.001) \).

When study groups were individually assessed, the vocal characteristics, creaky, pharyngealized and whispery, emerge from listeners’ voice ratings. The participants with RHD produced idiomatic utterances with creaky voice quality more often compared to literal utterances in both elicitation \( (t(125) = 3.33; p = 0.001) \) and repetition \( (t(125) = 2.73; p = 0.007) \) tasks. Creaky voice quality was also associated with LHD participants’ production of repeated idiomatic utterances compared to repeated literal utterances \( (t(125) = 2.74; p = 0.007) \) and HC’s production of elicited \( (t(125) = 3.5; p = 0.001) \) and repeated \( (t(125) = 2.98; p = 0.004) \) idiomatic utterances compared to literal counterparts.

Pharyngealized voice quality was judged to be associated with repeated idiomatic utterances compared to repeated literal utterances \( (t(125) = 4.03; p < 0.001) \) in RHD group. Idiomatic utterances produced in elicited \( (t(125) = 5.16; p < 0.001) \) and repeated modes \( (t(125) = 5.8; p < 0.001) \) by LHD participants were also judged to be pharyngealized compared to literal counterparts.

Finally, whispery was perceived in idiomatic utterances in elicited \( (t(125) = 4.37; p < 0.001) \) and repeated \( (t(125) = 3.1; p = 0.002) \) modes spoken by LHD speakers.

Discussion

The purpose of this study was to investigate listeners’ performance in identifying prosodic contrasts used to distinguish idiomatic from literal meanings of ambiguous sentences, spoken by persons with LHD or RHD and HC participants. Listeners successfully identified the intended meanings of sentences produced by HC in both elicited and repeated modes. The finding of native listeners’ ability to identify subtle prosodic cues is in accordance with previous studies, in which listeners successfully discriminated between idiomatic and literal utterances when speakers purposely conveyed the contrasting meanings (Belanger, Baum, & Titone, 2009; Sidtis & Van Lancker Sidtis, 2003; Van Lancker & Canter, 1981; Van Lancker Sidtis, 2003; Yang, Ahn, & Van Lancker Sidtis, 2015).

Utterances produced in either idiomatic or literal mode by participants with brain damage were less successfully identified by native listeners, with significantly decreased identification of both idiomatic and literal utterances produced by RHD. Based on the observation that native listeners have more difficulty identifying both kinds of utterances produced by individuals with RHD compared to ones produced by the LHD group, it can be concluded that RHD negatively affects the production of auditory and acoustic cues sufficient for contrasting idiomatic from literal utterances. This supports a role of the right hemisphere in the processing of prosodic cues (Bradvik et al., 1991; Bryan, 1989; Jackson, Taylor, Holmes, & Walshe, 1931; Weintraub, Mesulam, & Kramer 1981).
Speech obtained from brain-damaged individuals using two speech tasks (elicitation vs. repetition) yielded consistently different performance measures by listeners. Utterances produced by LHD and RHD during the elicitation tasks were identified significantly less successfully than those in the repetition tasks. Compared to other groups, idiomatic utterances produced in the elicited mode by the RHD group were identified less successfully by listeners. The findings that the intended meanings of idiomatic expressions produced by individuals with RHD during the elicitation tasks were less successfully identified by native listeners are also possibly due to impaired pragmatic abilities in RHD, which may extend to idiomatic expressions (Sidtis et al., 2009; Van Lancker Sidtis & Postman, 2006). These task-dependent differences support previous findings that phonatory and articulatory parameters of speech are affected by speech tasks (Canter & Van Lancker, 1985; Duffy, 1995; Kempler & Van Lancker, 2002; Kent, Kent, Rosenbek, Vorperian, & Weismer, 1997; Schalling & Hartelius, 2004; Van Lancker Sidtis, Kempler, Jackson, & Metter, 2010). The effect of task on speech production in brain-damaged participants may be attributable to the lack of an external model in spontaneous productions versus the presence of an external model in repetition. Individuals with brain damage may lack sufficiently effective internal models guiding motor acts associated with speech. These results highlight the importance of examining speech under different tasks.

Correct responses in the listeners’ identification tasks were highly correlated with listeners’ goodness ratings of utterances. The higher the goodness rating on the utterance, the more likely listeners were to identify the intended meaning of the utterance successfully. This result suggests that listeners have an internal model of the canonical forms of idioms, as well as how prosodic characteristics distinguish literal and idiomatic meanings in ambiguous sentences, which further underscores the status of these two sentence types in native competence. These findings agree with a perception-by-listeners study by Rammell, Pisoni and Van Lancker Sidtis (2016), in which formulaic expressions presented in noise were significantly more intelligible than matched grammatical expressions under the same listening conditions. This finding also implies storage in memory.

There was some evidence that idiomatic utterances differ from literal utterances in terms of voice quality. Perceptual ratings of vocal quality suggested that nasal and creaky voice qualities appear to be associated with idiomatic utterances. This result is compatible with findings from Van Lancker et al. (1981), who reported that nasal quality was also impressionistically notated for some of the US idiomatic utterances. Traditionally, nasal voice quality is associated with sarcasm, and idiomatic utterances tend to carry ironic nuances when compared to literal counterparts. Vocal qualities differentiating between idiomatic and literal utterances were not notably associated with hemispheric side of brain damage.

Taken together, the results indicated that healthy listeners can distinguish ambiguous sentences from prosodic information alone and that RHD negatively affects speech productions associated with idiomatic expressions. Differences in performance between two speech tasks, whereby prosodic productions were better in repetition than in elicitation, indicate a motor ability to produce the speech exemplars that persists even when spontaneous production is compromised. An explanation for preserved ability in repetition as in persons with unilateral brain damaged compared with elicited production
remains to be developed. Improved performance in repetition has been seen in other neurogenic speech disorders (Van Lancker Sidtis, Rogers et al., 2010). Studies comparing rote repetition (repeating an utterance produced by the examiner) with pragmatic (repetition in the course of a conversation) in persons with brain damage have received some recent attention (Oelschlaeger & Damico, 1998; Van Lancker Sidtis & Wolf, 2015; Wolf, Van Lancer Sidtis, & Sidtis, 2014). Superior performance during rote repetition is an important factor which deserves further research.

These findings also support the proposal that formulaic expressions differ from novel expressions in important speech characteristics and in native speaker’s knowledge. Formulaic expressions are unitary in form with distinctive prosodic features. Further, speakers in a language community know them, in that they are stored in memory with their auditory/acoustic and meaning characteristics. The literal counterparts can be distinguished from the idiomatic versions by native speakers using known features of the expressions, and the ability to communicate these auditory/acoustic characteristics is impaired or distorted in different ways by brain damage. There is indirect support from these results for the dual process model of language (Wray & Perkins, 2000), which envisions separate neural processing systems for formulaic and literal language in production and comprehension (Van Lancker Sidtis, 2008). These findings support previous reports of prosodic information serving to signal idiomatic meanings known to speakers in a community, the dual process model of language, a significant role of speech task in measures of speech production, and a right-hemisphere involvement in formulaic language.

Acknowledgments

We acknowledge the contributions to this study of John J. Sidtis and the Brain and Behavior Laboratory, Geriatrics Division, Nathan Kline Institute. We appreciate comments from Dr. Susannah Levi, Dr. Christine Reuterskiöld, Dr. Harriet Klein and Dr. Gerald Voelbel from New York University. Michele Burgevin provided editorial and graphic assistance. We would especially like to thank study participants at Myungji Choonhey Rehabilitation Hospital in Seoul, Korea.

Declaration of interest

The authors report no conflict of interest.

References


## Appendix 1

<table>
<thead>
<tr>
<th>Korean</th>
<th>Intended meaning</th>
<th>Situational/linguistic context</th>
</tr>
</thead>
<tbody>
<tr>
<td>그 아이는 이르를 갔어.</td>
<td>Idiomatic (The kid was so angry.)</td>
<td>그 아이는 자기가 좋아하는 장난감을 빼앗겨 화가 몽고서 했다. (The kid was so angry because his favourite toy was taken.)</td>
</tr>
<tr>
<td>1.</td>
<td>Literal (The kid ground his teeth.)</td>
<td>그 아이는 잠자는 동안 이르를 갔어. (The kid ground his teeth while he was sleeping.)</td>
</tr>
<tr>
<td>그 사람이 입을 다물었어.</td>
<td>Idiomatic (He kept silent.)</td>
<td>그 사람은 그 사건에 대해 아무 말도 하지 않았다. (During the police investigation, he did not say anything.)</td>
</tr>
<tr>
<td>2.</td>
<td>Literal (He had his mouth closed.)</td>
<td>그 사람은 치과수술 후 입을 벌이지 않았다. (After the oral surgery, he could not open his mouth.)</td>
</tr>
<tr>
<td>그 사람이 별집을 건드렸어.</td>
<td>Idiomatic (He caused a big problem.)</td>
<td>그 사람은 상사의 혐의를 인터넷에 올렸는데, 상사가 그것을 보았다. (He posted his negative comments about his boss on his personal website, which his boss saw.)</td>
</tr>
<tr>
<td>3.</td>
<td>Literal (He stirred up a beehive.)</td>
<td>그 사람이 산에 벌이 많아서 그 사람이 별집을 만지지 않도록 주의했다. (He tried not to touch a beehive when he did hiking.)</td>
</tr>
<tr>
<td>그 사람은 발을 굴려.</td>
<td>Idiomatic (He was nervous.)</td>
<td>그 사람은 마지막 차시간에 맞추지 못할까봐 긴장했다. (He was so nervous since he almost missed the train.)</td>
</tr>
<tr>
<td>4.</td>
<td>Literal (He stomped his feet.)</td>
<td>그 사람은 총추는 기술을 배웠다. (He stomped his feet during dancing.)</td>
</tr>
<tr>
<td>그 사람은 무릎을 꿇었어.</td>
<td>Idiomatic (He surrendered.)</td>
<td>그 사람은 할 수 있는것을 다해봤지만 결국엔 포기했다. (He ended up giving up his dream.)</td>
</tr>
<tr>
<td>5.</td>
<td>Literal (He kneeled down.)</td>
<td>그 사람은 아침마다 부모님께 절을 한다. (He always started his day by performing one deep traditional bow to their parents.)</td>
</tr>
<tr>
<td>그 사람은 콩밥을 먹었어.</td>
<td>Idiomatic (He served a prison term.)</td>
<td>그 사람은 감옥에 갔다 혼적이 있다. (He was charged with bank fraud and went to jail.)</td>
</tr>
<tr>
<td>6.</td>
<td>Literal (He ate bean-mixed rice.)</td>
<td>그 사람은 오늘 아침 콩밥에 미역국을 먹었다. (He had bean-mixed rice and seaweed soup for breakfast.)</td>
</tr>
</tbody>
</table>