

Three Types of Near-Synonymy and Rare Collocations: Selected Results from a Cognitive, Corpus-Based Study of Nominal Lexical Substitution in Korean, Japanese, and English Idioms

CAREY BENOM

Seinan Gakuin University

YOUNG MIN OH

Ritsumeikan Asia Pacific University

1 Introduction

In this paper, we present and discuss selected findings from our ongoing study of idiomatic creativity, focusing on a specific type of creative variation that we categorize as ‘lexical substitution’. We use this term to refer to the creative use of an idiom in which a lexeme in the canonical form of the idiom has been replaced with another lexeme or phrase. For instance, instead of the canonical *let the cat out of the bag* ‘reveal the secret’, consider *let the kittens out of the bag*, in which *kittens* has been substituted for the canonical *cat*. Further narrowing our scope, we will focus here on noun substitution, as exemplified by *kittens* for *cat*.

We adopt an inductive approach in our study of the range and patterns of nominal lexical substitution found in creative uses of fifteen idioms each in Korean, Japanese, and English. Our purposes are to reveal the relevant semantic categories; to attempt to motivate them; to refine and expand on previous analyses; and to shed further light on the cognitive and linguistic strategies employed for the creative manipulation of idioms (see Benom & Oh 2020, 2023, Oh 2022, Benom

Japanese/Korean Linguistics 31.

Edited by William Giang, Lucien Brown, Shimako Iwasaki, Satoshi Nambu, and Daniel Pieper.

Copyright © 2024, Carey Benom and Young Min Oh

2023). Our findings reveal robust patterns, which illustrate speakers' strategies as well as motivations for altering idioms in various ways.

Here, we will report on two key phenomena from among the wide variety of intriguing patterns observable in our data which show nominal substitution: multiple categories of near-synonymy, based on different senses of the replaced noun, and rare collocations, which are nevertheless immediately comprehensible. These rare combinations of words can emerge, for instance, when a substituted noun which preserves the basic figurative meaning of the canonical, missing noun, appears together with the remainder of the idiom, as we will detail below.

2 Background

Idioms are highly idiosyncratic expressions whose meanings are not straightforwardly derivable from the individual meanings of their components. This is because idioms do not adhere strictly to compositionality—although compositionality exists on a spectrum rather than as an absolute distinction (Wulff 2008: Chapter 3).

Idioms are typically defined as 'conventional, non-compositional multi-word expressions'. This definition covers everything from phrasal verbs such as *get up* to sentence-long proverbs such as *don't count your chickens before they hatch*. We have no objections to such a definition, but to consider idioms as constructions, we might define them instead as 'substantive multi-lexemic conventional symbolic units'. Specifically, 'substantive' refers to a construction that is partially or wholly 'filled in' (Croft & Cruse 2004); 'conventional' means that it is generally used in a speech community; 'symbolic' refers to the pairing of form and function; and a 'unit' is defined as an entrenched routine (see Langacker 1987, Croft 2001).

While idioms are often characterized as 'fixed' or 'frozen' (e.g. Miyaji 1982, Gibbs 2007, Ishida 2015), research—including our own and that of others (e.g. Moon 1998, Oh 2022)—demonstrates that certain idiomatic variations occur even more frequently than their canonical forms. Recently, a broad range of idiom uses displaying impressive creativity has been the subject of an increasing number of investigations based on corpus data (Moon 1998, Langlotz 2006, Wulff 2008, Tsuchiya 2013, Ishida 2014, Oh 2018, 2019, 2020a, 2020b, 2022, Benom & Oh 2020, 2023, Oh & Benom 2021, Benom 2023). It is within the context of these studies on idiomatic creativity that the present paper emerges.

Speakers regularly use idioms in creative ways, and because the full breadth of possibility of this creativity, as well as the complete list of factors motivating idiomatic creativity, remain somewhat elusive, further investigation has the potential to provide valuable insights for the study of linguistic theory, language description, phraseology, formulaic language, collocations, proverbs, and second language teaching and learning.

3 Methods

We take an inductive, data-driven approach. Our working definition of idiomatic creativity relies on deviations from the standard or canonical form, which we established by consulting the widest possible range of idiom dictionaries for each language, applying a majority-rule approach to resolve discrepancies.

We used large, comparable, web-based corpora and systematically extracted, manually examined, and coded 8,207 total creative uses with lexical substitution ($K = 3,149$; $J = 2,059$; $E =$

2,999), and it is from here that the data on noun substitution are drawn. For Korean, we used the koTenTen18 (1.7 billion words); for Japanese, the jaTenTen11 (8.4 billion words); and for English, the enTenTen15 (13 billion words).

We attempted to extract all the idiomatic variation from each corpus by searching for the key elements of an idiom (e.g. *let*, *cat*, *out*, and *bag* in the idiom *let the cat out of the bag*) two at a time (e.g. *let* and *cat*, *let* and *out*, *let* and *bag*, etc.) within a span of ten words and manually filtering the results. This method required a substantial time investment but resulted in more valid, reliable, and complete data than that obtainable through most alternative approaches. For both Korean and Japanese, we analyzed all data retrieved, but due to the larger size of the English corpus and the abundance of certain results, we capped our analysis of English data at the first 500 results whenever the total exceeded 1,000. After coding the data, we isolated all instances of substitution and analyzed the semantic relationship between the canonical and substituted lexemes, as well as the semantic contribution of the substituted lexeme in its idiomatic context.

Relying on web data is not ideal for several reasons, the most critical of which is that language use online differs significantly from spoken language. As a result, our findings cannot be assumed to apply more broadly to spoken language in general. Nevertheless, the sheer size of these corpora enabled us to gather a substantial number of relevant examples for analysis. Despite this, data sparsity remained an issue for certain idioms, including for English, even with our extensive datasets.¹

To measure the strength of collocation of two terms, as a way of supporting our claims that certain word combinations are unusual, we elicited a statistical value called a logDice score (Rychlý 2008).² This metric evaluates the cooccurrence of two terms, X and Y, by considering their combined frequency relative to the individual frequencies of X and Y. The theoretical maximum logDice score is 14, which would indicate that all instances of X appear with Y and vice versa, though in practice, even the highest scores tend to remain below 10. In terms of interpretation, a +1 increase in logDice corresponds to a doubling of collocation frequency, while a +7 increase suggests an approximately 100-fold increase. Low or negative logDice scores indicate that there is no statistical significance to the collocation, and the lower the score, the more it suggests that there is a semantic clash, for example, or that the cooccurrence is random. We set a logDice value of 3 as the cutoff to reveal a meaningful collocation attraction.

4 Results and Discussion

We will first describe and illustrate the different types of near-synonymous substitutions we observed. Following this, we will present and discuss some of the rare collocations present in the data.

¹ One result of this data sparsity is that the number of idioms in each language which we might consider for analysis depends on the type of variation studied, and therefore the total number of idioms we cite in previous work (e.g. Benom & Oh 2023) differs from that given here.

² After we compiled our data, the Sketch Engine website released a newer English web corpus: the enTenTen21, which contains 54 billion words. Due to this substantial increase in size, and the resulting increase in confidence in the outcome of statistical tests applied to the data from this corpus, we decided to run such tests for rare collocations in English, described below, in this newer corpus, rather than the enTenTen15 from which the data came, since we are claiming that these collocations are not only or specifically rare in the corpus in which the data were found but in general.

4.1 Three Types of Near-Synonymy

Across all three languages, for all idioms, noun substitutes were frequently rough synonyms of the replaced term.³ Our analysis reveals three distinct types of substitution based on such near-synonymy in our data. These variations enabled speakers to fine-tune idiomatic expressions to align more precisely with their communicative intent. The three types differ based on which aspects of the original lexeme's meaning are activated in the modified variant.

4.1.1 Near-Synonymy Type 1: Literal

This type of near-synonymy invokes the literal meaning of the replaced term, introducing nuance while maintaining the idiom's overall semantic structure. Consider the corpus data behind an already familiar English example, in which *kittens* has been substituted for *cat* in the idiom *let the cat out of the bag*:

- 1) *I kept a few chapters for myself, can't let all the kittens out of the bag at once*

Here, the use of *kittens* emphasizes aspects such as innocence, smallness, and multiplicity. Such substitution demonstrates how speakers can introduce subtle variations that play on the literal interpretation of the replaced noun. It is used to carve out a more detailed semantic representation and can help with reference. A similar phenomenon was observed in noun substitutions for the term *pwul* 'fire' in the Korean idiom presented below in its canonical form:

- 2) *paltung-ey pwul-i tteleci-ta*
 top of the feet-LOC⁴ fire-NOM fall-DEC
 Lit. 'a fire falls on top of one's foot'
 Fig. 'an urgent issue suddenly strikes'

Among the creative uses for this idiom, we found *pwul* 'fire' (fig. 'urgent issue') replaced with elements such as *pwulttong* 'spark', *pwultengi* 'fireball', and *pwulssi* 'embers'. These noun substitutions subtly refine the semantics of the idiom by enhancing the sense of urgency (*pwultengi* 'fireball') or by invoking a connection to a potential future danger (*pwulttong* 'spark', *pwulssi* 'embers'). Such examples showcase language users' ability to creatively manipulate idiomatic language to add nuance while maintaining a clear connection to the original expression.

4.1.2 Near-Synonymy Type 2: Figurative

This category of near-synonymy invokes the figurative meaning of the replaced term. It results in a more explicit expression of the underlying concept, facilitating reference by directly assigning a referent to its role in the idiom frame (Fillmore 2006), and reducing potential ambiguity. Observe the data below:

- 3) *Well I have finally decided to let a secret out of the bag*

³ This is true as well for substitutes for other parts of speech, but this issue is beyond our present scope

⁴ Abbreviations used in this paper include: ACC = accusative; DEC = declarative; LOC = locative; NF = non-final; NOM = nominative; PFV = perfective; POL = polite; PRES = present; REL = relativizer; TR = transitive.

With this example, there is no guesswork necessary, as the creative variant directly invokes the figurative meaning of the replaced lexeme (*cat*). A similar situation holds with certain substitutions for the Korean idiom in (2) above, including in the creative use we present below:

- 4) *paltung-ey* *tteleci-n* *mwuncey-nikka-yo*
 the top of the feet-LOC fall-PFV.REL problem-because-POL
 Lit. ‘because it is a problem that has fallen on the top of the feet’
 Fig. ‘because it is an urgent problem that needs immediate attention’

In this use, *pwul* ‘fire’ (fig. ‘urgent issue’) is replaced with *mwuncey* ‘problem’, based on the figurative meaning of the missing term in the idiomatic context. This type of substitution shows how speakers can creatively adapt idioms to make their figurative meanings more transparent. This can be a perfect solution to the problem of the complexity of reference inherent in creative idiom use (Benom & Oh 2020).

4.1.3 Near-Synonymy Type 3: Extended

This third type of near-synonymy invokes an extended sense of the replaced term. It appears to be rare, as we only found this phenomenon with one idiom—the English *penny wise and pound foolish* ‘to be careful with small money (fig. “smaller issues”) but to be wasteful or mistaken about larger money (fig. “larger issues”)’. Though there were multiple instances in the data, they all relied on the same lexeme—*pound*. Among the substitutes for *pound*, we found instances of Type 1, literal near-synonymy such as *dollar* and *euro*, as well as instances of Type 2, figurative near-synonymy such as *life* and *justice*, in which speakers explicitly noted the larger issues that they believed were being mishandled, but we also found additional instances of near-synonymy, including *454 grams*, *kilo*, and *ton* (this last one not a very near synonym), which employ the sense of *pound* as a unit of weight rather than a unit of currency.

These substitutions likely stem from wordplay and demonstrate a high degree of linguistic creativity. They show how speakers can exploit multiple semantic dimensions of a word within idiomatic contexts. They also demonstrate speakers’ ability to creatively manipulate the polysemy of individual lexemes within larger idiomatic structures, all while maintaining relevance to the discourse, showcasing an impressive level of cognitive flexibility. Due to the limitations of our data, however, whether other lexemes in other idioms are employed in this way remains open for future research.

4.2 Rare Collocations

Beyond near-synonymy, our data revealed other types of noun substitution. A full enumeration is beyond our present scope,⁵ but here, we will describe the most frequent type of noun substitution and demonstrate how it can produce unusual yet immediately interpretable word combinations, thanks to the rich semantic potential of idioms and the complexity of referentiality involved with their use (Benom & Oh 2020). Such rare collocations were the result of speakers creatively map-

⁵ Though we did so, somewhat preliminarily, in earlier work such as Benom (2023) and Oh (2022); however, analysis of additional data, related to our present effort, warrants a more extensive discussion and an updated taxonomy, which we save for future research.

ping referents to the roles of the idiom's frame to aid in reference (Benom & Oh 2020, Fillmore 2006).

As for the type of noun substitution we wish to describe: The noun that is substituted into the idiom refers to a discourse-relevant instance of the category defined by the figurative meaning of the replaced noun in the context of the idiom. We refer to the general creative strategy this reveals as 'exemplification' (and we will note that, in addition to the category of noun substitution we will discuss here, exemplification motivates other types of idiomatic creativity, as we will demonstrate below). An English example of such a noun substitution which is motivated by exemplification, resulting in a rare collocation, is given below:

- 5) *We need to be serious and design a system that does more than count our teachers before they hatch*

In this example, *chickens* (fig. 'potential future good things') in the canonical idiom has been replaced with *teachers*—an instance of the category of potential future good things, which is embedded in the larger discourse context. Due to this use of exemplification, *teachers* are said to (potentially) *hatch*—a combination of lexemes that is immediately comprehensible, given the idiomatic context, despite being very rare indeed, as shown by their strikingly low logDice score of approximately -11.07. This logDice score is the result of *teachers* appearing 3,981,828 times in the enTenTen21 corpus, *hatch* appearing 246,855 times, and the combination appearing just 33 times (within a span of 10 words of one another).

While the combination of *teachers* and *hatch* is rare, such interesting word combinations in the data are not. Similar statistical results were observed with other substitutions for *chickens* in the same idiom, such as *Oscars* in *don't count your Oscars before they hatch*. Again, *Oscars* refers to an instance of the potentially good thing (an award) on which one should not rely, a result of exemplification. The logDice of *Oscars* and *hatch* is approximately -1.49, suggesting no attraction whatsoever.

Let us return to the idea that exemplification does not exclusively apply to noun substitution. This strategy can also motivate what we code as 'extension', or the addition of lexemes to the canonical idiom form (e.g. Oh 2022). Such extensions, too, can result in rare collocations. For example, consider both *commodities* (which we coded as an extension) and *investments* (coded as a substitution) in the use below.

- 6) *Australia therefore should not be counting its commodities investments before they hatch*⁶

In this case, both *commodities* and *investments* should be considered the result of a single instance of the speaker applying the strategy of exemplification, since *commodities investments* is an instance of the category defined by the figurative meaning of the replaced *chickens*.

Furthermore, when paired with *hatch*, the substitution *investments* and the extension *commodities* showed logDice scores of approximately -11.08 and -11.5, respectively, strongly suggesting a lack of any attraction whatsoever and attesting to the fact that the word combinations are extremely rare.

⁶ See Benom and Oh (2020) for a discussion of this example.

Rare collocations are produced based on contextual relevance and require both categorical alignment and cognitive flexibility in production and comprehension. By selecting discourse activated instances that fit the category defined by the idiomatic meaning of the replaced noun, speakers engage in a complex cognitive task. They must maintain the overall figurative meaning of the idiom while simultaneously introducing novel elements that are relevant to the current discourse context. This process showcases the dynamic interplay between fixed idiomatic structures and creative language use, highlighting the flexibility of human cognition and of creativity in linguistic expression.

Exemplification, which we have so far demonstrated with *Oscars*, *teachers*, and *investments* as substitutes for *chickens*, and *commodities* as an extension, is the most frequent cause of such unusual word combinations, but other routes of idiomatic creativity can also lead to rare but immediately comprehensible combinations of lexemes. We will return to this issue below.

As our next example of exemplification leading to rare word combinations, we first offer the Japanese idiom below in its canonical form:

- 7) *yabu wo tsutsui-te hebi wo das-u*
 stump ACC poke-NF snake ACC out.TR-PRES
 Lit. ‘to poke a tree stump and cause snakes to appear’
 Fig. ‘to stir up trouble, to not let sleeping dogs lie’

Substitutes for *hebi* ‘snake’ (fig. ‘dangerous thing’) included *seshiumu* ‘cesium’ and *beigun kichi no kakudai ya jyōho* ‘the expansion and concessions of U.S. military bases’, (elements of) which infrequently collocate with *yabu* ‘stump’. For instance, *yabu* ‘stump’ plus *kichi* ‘base’ results in a logDice value of -0.23, while *yabu* ‘stump’ plus *seshiumu* ‘cesium’ gives us a value of -0.21, supporting our claims of statistical independence.

Yet, once again, thanks to the context of the idiom, the unusual word combinations are easy to understand. Such substitutions demonstrate, as we also saw above, how speakers actively map referents to the roles within an idiom’s semantic frame. This mapping process serves to aid in reference and contextualization of the noncanonical idiomatic expression.

As a final example of noun substitution motivated by exemplification, consider first the Japanese idiom presented in its canonical form below:

- 8) *tora-nu tanuki no kawa-zanyō*
 capture-NEG raccoon-dog GEN skin-calculation
 Lit. ‘to count the skins of racoon dogs that you have not yet caught’
 Fig. ‘don’t count your chickens before they hatch’

Our data include a creative use in which *tanuki* ‘raccoon dog’ is replaced by *kyūryō* ‘salary’, resulting in a literal meaning translatable as ‘to count the skin(s) of salary that you have not yet caught’. Here, *kyūryō* ‘salary’ exemplifies the category of ‘anticipated but not yet realized gains’, aligning with the figurative meaning of the absent canonical term *tanuki* ‘raccoon dog’. Despite the unusual collocation (*kyūryō* ‘salary’ + *kawa* ‘skin’; logDice = -2.19), the expression is immediately comprehensible due to the idiomatic framework. In fact, the task of determining reference is drastically simplified by the substitution.

At this point, having demonstrated some results of the application of the strategy of exemplification in English and Japanese, and having shown how it can result in rare collocations via the use of both substitution and extension, we will return to the idea that rare collocations can result from other phenomena. In fact, we will cite some examples from above.

First, we will return to the Korean example in (4) above, repeated below.

- 9) *paltung-ey* *tteleci-n* *mwuncey-nikka-yo*
 the top of the feet-LOC fall- PFV.REL problem-because-POL
 Lit. ‘because it is a problem that has fallen on the top of the feet’
 Fig. ‘because it is an urgent problem that needs immediate attention’

This creative use, as mentioned, is based on the canonical idiom seen in (2) above. As described, it is an instance of Type 2 synonymy, based on the figurative meaning of the replaced term. In this example, *tteleci-n* ‘fall’ combining with *mwuncey* ‘problem’ results in a logDice value of 0.81, which is far below the cutoff value of 3, suggesting that the two lexemes lack any attraction for one another.

Some of our examples of Type 3 near-synonymy, presented above in Section 4.1.3, involve rare collocations as well. The logDice value for the word pair *penny* and *kilo*, found in the creative use *penny wise and kilo foolish*, is approximately -9.16. That for *penny* and *ton*, found in *penny wise and ton foolish*, is approximately 1.74. This is to say that, in addition to the type of noun substitution motivated by exemplification, as well as similarly motivated extension, both Type 2 and Type 3 near-synonymy can result in rare collocations. Such cases highlight how idiomatic creativity can simultaneously rely on categorical alignment, semantic flexibility, and discourse-driven adaptation.

5 Conclusions and Future Directions

In this paper, we highlighted two key findings from our study of nominal substitution as a form of idiomatic creativity. First, we demonstrated that speakers employ different types of near-synonymy based on their communicative goals, showing:

- 1) Near-synonymy that invokes the literal meaning of the replaced term, such as *let the kittens out of the bag*, in which speakers enhance idiomatic expressions by integrating specific details related to the literal sense of the replaced noun, adding nuance and clarifying reference.
- 2) Near-synonymy that explicitly invokes the figurative meaning of the replaced term, such as *I’ve let my secret out of the bag*, in which the figurative layer is removed and the use directly signals the intended meaning, enhancing clarity.
- 3) Near-synonymy that invokes an extended sense of the replaced term, such as *penny wise and kilo foolish*, introducing a conceptual twist to the idiom’s interpretation by playing with polysemy. This phenomenon was limited to the term *pound* in the idiom *penny wise and pound foolish* among those we studied.

Next, we described rare collocations, initially showing that they can result from a certain, frequent type of noun substitution in which the substituted noun exemplifies the category defined by

the canonical, missing term, in the idiomatic context. We claimed that this type of substitution is motivated by exemplification. We argued that the strategy of exemplification also motivates other types of creativity, such as extension, which can result in rare collocations as well. Finally, we showed examples of two types of synonymy resulting in rare collocations as well, revealing that the phenomenon of creative uses of idioms resulting in unusual word combinations has multiple, distinct causes.

Beyond the types of creativity described here, our data revealed additional forms of noun substitution, including the use of antonyms, homonyms, generalizations, various kinds of wordplay, and analogies. A full treatment of nominal substitution in creative uses of idioms awaits future research. Finally, we will mention that verb substitutes did not behave like noun substitutes, as they altered certain semantic features of the canonical verb (causation, intentionality, dynamicity, and polarity) and the number of roles in the idiom frame, as well as the mapping of those roles to syntax, but they did not result in rare collocations, at least in our data. We will save the issue of verbal substitution for future research.

References

- Benom, C. 2023. Argentina, Eat your Cows Out! Lexical Substitution in English and Japanese Heart Idioms. *Embodiment in Cross-Linguistic Studies*, eds. J. B. Kóczy and K. Sipocz, 369–91. Leiden/Boston: Brill.
- Benom, C. and Y. M. Oh. 2020. Don't Count your Commodities Investments before they Hatch: Idiomatic Creativity, Concretization, and the Comprehensibility of "Impossible" Collocations. Paper presented at *The 56th Linguistics Colloquium ("Multilingualism, Translation and Cognition")*, online, November 26, 2020.
- Benom, C. and Y. M. Oh. 2023. 'Keepwords', the Limits of Creativity, and the Notion of the "Core" of an Idiom. *Japanese/Korean Linguistics 30*, eds. S. Williamson, A. A. Babayode-Lawal, L. Bosman, N. Chan, S. Cho, I. Fong, and K. Holubowsky, 223–37. Stanford, CA: CSLI Publications.
- Croft, W. and D. A. Cruse. 2004. *Cognitive Linguistics*. Cambridge: Cambridge University Press.
- Fillmore, C. J. 2006. Frame semantics. *Cognitive linguistics: Basic readings*, ed. D. Geeraerts, 373–400. Berlin/New York: Mouton de Gruyter
- Gibbs, R. W. Jr. 2007. Idioms and formulaic language. *The Oxford Handbook of Cognitive Linguistics*, eds. D. Geeraerts and H. Cuyckens, 697–725. New York: Oxford University Press.
- Ishida, P. 2014. Nihongokopasu-ni-mirareru-kanyōku-no-henkakanosei (Variability of Idioms in Japanese Corpora). *Kenshūsha Web Magazine Lingua* 13.
- Ishida, P. 2015. *Gengo-kara-mita-nihongo-to-eigo-no-kanyōku* (Idioms in Japanese and English from a Linguistic Perspective). Tokyo: Kaitakusha.
- Langlotz, A. 2006. *Idiomatic Creativity*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Moon, R. 1998. *Fixed Expressions and Idioms in English: A Corpus-based Approach*. New York: Oxford University Press.
- Miyaji, Y. 1982. *Kanyōku-no-imi-to-yōho* (The Meaning and Usage of Idioms). Tokyo: Meijishoin.
- Oh, Y. M. 2018. Adding Insult to Spilled Coke: A Corpus-based Study of Idiomatic Creativity. *Proceedings of the Japanese Cognitive Linguistics Society* 17: 94–106.
- Oh, Y. M. 2019. A Web-Corpus based approach to creative variation in several Korean idiomatic expressions. *Korean Culture Education Research* 4: 141–60.

- Oh, Y. M. 2020a. A Systematic, Corpus-Based Methodology for Studying Idiomatic Creativity in Korean and Japanese. *Japanese/Korean Linguistics* 27, ed. M. Barrie, 87–103. Stanford, CA: CSLI Publications.
- Oh, Y. M. 2020b. A Corpus-based Study of Idiomatic Creativity: A Case Study of “me kara uroko ga ochiru” and “oya no sune o kaziru”. *KLS Selected Papers 2*, ed. Kansai Linguistic Society, 103–18. Osaka: Kansai Linguistic Society.
- Oh, Y. M. 2022. *A Web Corpus Driven Study of Idioms and their Variations in English, Japanese, and Korean from a Cognitive Linguistic Perspective*. Doctoral dissertation, Kansai University.
- Oh, Y. M. and C. Benom. 2021. Contraction as Idiomatic Variation. *Japanese/Korean Linguistics* 28, eds. H. Jeon, P. Sells, Z. You, S. Kita, and J. Yeon, 361–75. Stanford, CA: CSLI Publications.
- Rychlý, P. 2008. A Lexicographer-Friendly Association Score. *Proceedings of Recent Advances in Slavonic Natural Language Processing*, eds. P. Sojka and A. Horák, 6–9. Brno: Masarykova Univerzita.
- Tsuchiya, T. 2013. *Teikeihyogen-o-kiban-to-shita-genga-no-sozosei: Kanyohyogen-to-kotowaza-no-kaku-choyoho-ni-kansuru-shakai-ninchiteki-kosatsu* (Language Creativity Based on Formulaic Language: Extensional Usage of Idioms and Proverbs from a Social Cognitive View). Doctoral dissertation, Kyoto University.
- Wulff, S. 2008. *Rethinking Idioms: A Usage-based Approach*. London/New York: Continuum.