

Varietal Difference in the Grammaticalization of *Keep V-ing*

Jonathan Jones and Michael Wolfman

Grammaticalization of *Keep V-ing*

Catenative Verb Constructions (CVCs)

- “A large class of constructions where a verb has a non-finite internal complement” (Huddleston et al. 2002)
- Part of an intermediate category between a prototypical auxiliary verb and prototypical full lexical verb that covers different semantic meanings such as modality or aspect (Quirk et al. 1985)
- Catenative verb constructions in English are a type of serial verb construction (Roberts 2011)

Iterative vs. Continuative Aspect

- Underwent two stages of grammaticalization (Santos 2017)
 - Full Lexical Item → Continuative Aspectual Marker
 - Continuative Aspectual Marker → Iterative Aspectual Marker
- Marks for two aspects (Santos 2017):
 - Iterative aspect when V-ing is characterized by a telic aktionsart
 - Continuative aspect when V-ing is characterized by an atelic aktionsart

Hypothesis

- **British English selects more for the iterative *keep V-ing* construction than American English.**

Methods

Corpora Used

- The British National Corpus (BNC)
- The Corpus of Contemporary American English (COCA)

Data Collection

- Queried searches to find *keep V-ing* constructions in the BNC & COCA
 - [lemma="keep"] [pos="v.g"] ([word!="by"] [pos!="n.*"]);
- Reduced the ~4,000 tokens (BNC) and ~33,000 tokens (COCA) to 3,000 samples from each corpus

Wilcoxon Paired Rank Test

- Ran the Wilcoxon paired rank test on the attestation rate of the top 100 V-ing to confirm the samples were representative of the corpora
- High P-Value is indicative of strong representation
 - BNC: 0.9319
 - COCA: 0.9892

Data Annotation

- Samples split evenly between annotators
- Annotated for aspect (C/I), tense, subject person, and V-ing
 - Any mistagged or passive instances were annotated accordingly and excluded from future analysis
 - Total BNC Samples: 2,891
 - Total COCA Samples: 2,970

Inter-Annotator Agreement

- A random set of 100 tokens was taken from each corpus and individually annotated for aspect to determine agreement using the Weighted Kappa Test
- Weighted Kappa Test Results:
 - Ambiguous aspect is in-between continuative and iterative, so a C/I to A pair was given an agreement weight of 0.5 rather than 0.
 - Expected agreement: 0.536
 - Actual agreement: 0.853
 - Kappa: 0.682 “Substantial Agreement” (Viera and Garrett 2005)

Significance Testing

- Annotated samples were then loaded into R and underwent Chi-squared testing to determine the effects of language variety, tense, subject person, and V-ing on the interpreted aspect of *keep*

Significance Testing

Variety of English

	COCA	BNC
Continuative	997 (33.6%)	788 (27.3%)
Iterative	1735 (58.4%)	1888 (65.3%)
Ambiguous	238 (8%)	215 (7.4%)
TOTAL:	2,970	2,891

X-Squared	Df	P-Value
31.041	2	1.818e-07

Tense

	PAST	PRESENT	FUTURE
Continuative	597 (25.2%)	1096 (33.3%)	92 (44.2%)
Iterative	1618 (68.4%)	1915 (58.3%)	90 (43.2%)
Ambiguous	152 (6.4%)	275 (8.4%)	26 (12.5%)
TOTAL:	2,367	3,286	208

X-Squared	Df	P-Value
91.022	4	2.2e-16

Subject Person

	1 st	2 nd	3 rd	N/A
Continuative	388 (28.8%)	386 (40.9%)	975 (27.8%)	36 (56.2%)
Iterative	858 (63.6%)	457 (48.5%)	2286 (65.2%)	22 (34.4%)
Ambiguous	103 (7.6%)	100 (10.6%)	244 (7%)	6 (9.4%)
TOTAL:	1,349	943	3,505	64

X-Squared	Df	P-Value
112.62	6	2.2e-16

Observations

- Variety of English, tense, and subject person all have a statistically relevant relationship with which aspect of *keep* is realized
- The V-ing has a statistically relevant connection; however, these Chi-Squared test results are not especially useful due to the extremely high X-squared and df values
 - X-squared = 5209.1, df = 2008, p-value < 2.2e-16
- Semantic categorization of these verbs would allow for more succinct data and fruitful analysis

Semantic Categorization

Difficulties

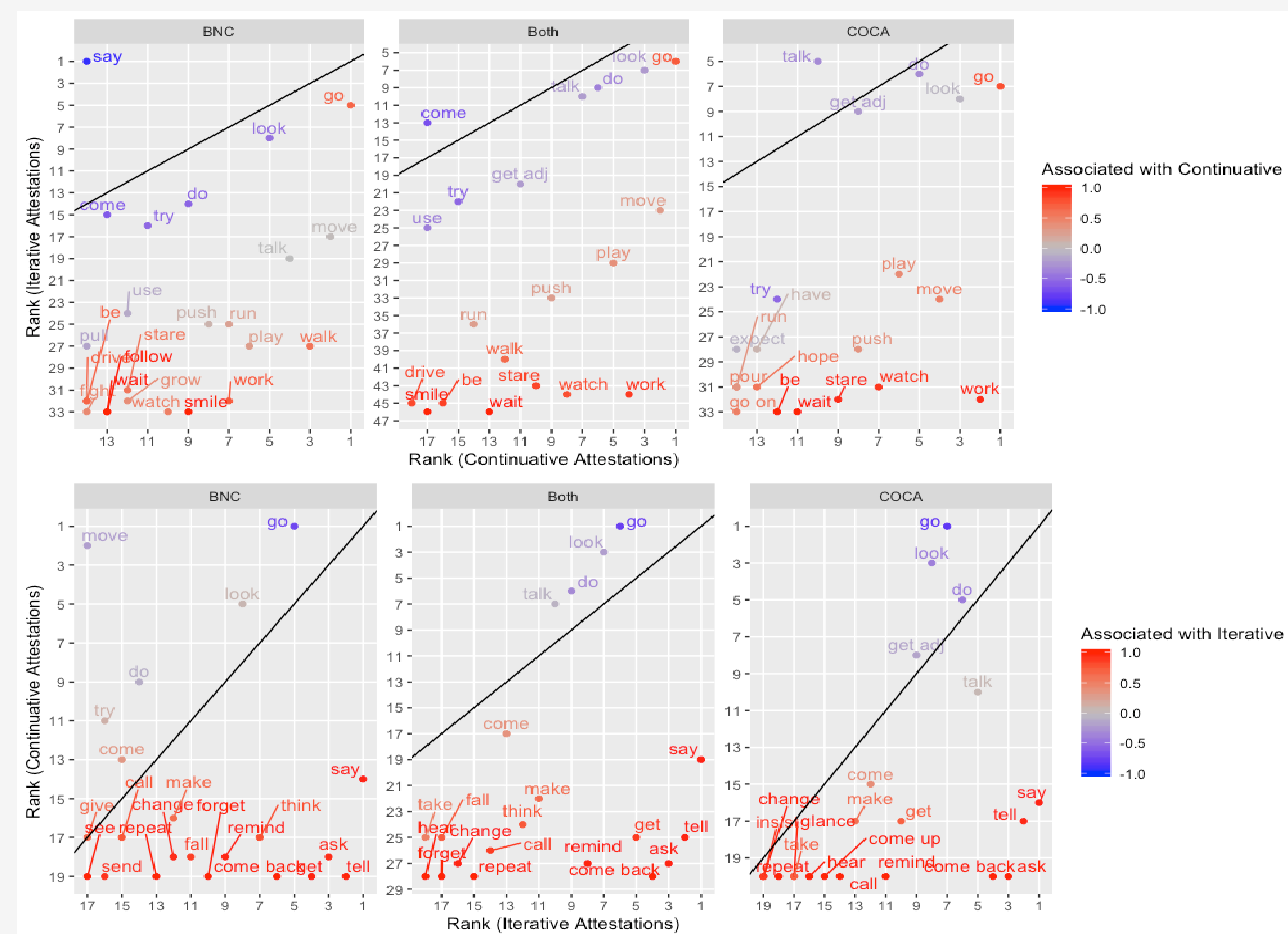
- Aktionsart annotation of V-ing would allow for the type of semantic categorization that is easily analyzed
- However, this process is hindered by a massive time and labor burden, as well as annotator subjectivity

Alternative

- As a preliminary substitute, association metrics were used with the V-ing rank of each aspect to create “aspect spaces”
- The “association” metric is the number of attestations of a verb with the given aspect minus attestations of the word in the other aspects, divided by total attestations.
- If the word ranks of each aspect were identical, they would match the given line (slope = 1)
- However, they do not, indicating a significant relationship between aspect and V-ing
- Verbs above the line have higher ranks in the aspect on the vertical axis than in the aspect on the horizontal axis

Observations

- The majority of these verbs fall under the line, indicating that continuative and iterative aspects are strongly associated with different sets of verbs.



Conclusions

Current Conclusions

- British English selects more for the iterative *keep V-ing* constructions than American English
 - It is possible that the two varieties of English are in different developmental stages of the grammaticalization process
- Grammatical elements such as tense and subject person play a significant role in selecting for iterative or continuative aspect
- Continuative and Iterative aspects have unique distributions, with little overlap, of the verbs with which they co-occur

Future Research

- Expand on semantic categorization of V-ing using a feature paradigm
- Explore the development and patterning of this type of CVC in other languages
 - Preliminary Survey of Welsh : *cadw* is emerging as a marker for both aspects, potentially through language contact, while continuative aspect uses *dal i* or *parhau* and iterative either uses *dal ati* or is indicated with an adverb.

References

- Davies, Mark. (2008-) The Corpus of Contemporary American English (COCA): 560 million words 1990-present. Available online at <https://www.english-corpora.org/coca/>.
- Huddleston, Rodney & Geoffrey K. Pullum. 2002. The Cambridge grammar of the English language. Cambridge: CUP.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech & Jan Svartvik. 1985. A comprehensive grammar of the English language. London: Longman.
- Roberts, John R. “Serial Verbs in English: An RRG Analysis of Catenative Verb Constructions.” *Functions of Language*, vol. 19, no. 2, 2012, pp. 201–234., doi:10.1075/fo1.19.2.03rob.
- Santos, Sara Sánchez-Molina. The grammaticalization of *Keep V-ing*. Leiden University Repository, 2017.
- The British National Corpus, version 3 (BNC XML Edition). 2007. Distributed by Bodleian Libraries, University of Oxford, on behalf of the BNC Consortium. URL: <http://www.natcorp.ox.ac.uk/>
- Viera AJ, Garrett JM. Understanding Interobserver Agreement: The Kappa Statistic. *Fam Med* 2005;37(5):360-363