Using the CORE Corpus and a multivariate analysis to (re)examine the impact of register and structural factors on *that*/zero complementizer variation in five mental state verbs

Christopher Shank (Bangor University, UK) and Koen Plevoets
(FLAMES – Lstat, Belgium)

This paper examines the distribution of *that*/zero complementation alternation patterns, and related claims, concerning the proposed roles that register (specifically type of register) and concomitant structural factors/variables within the matrix and complement clauses play as predictors of the presence of the zero form, within five mental state verbs (MSVs) viz. *think*, *believe*, *feel*, *know* and *understand*.

I.   I think *that*/zero I want to have a baby. (Register: Narrative, CORE)
II. He believes *that*/zero governments must not mandate prayer. (Register: Opinion, CORE)
III. They feel *that*/zero their job is done. (Register: Personal Blog, CORE)
IV. She knows *that*/zero it's not an option.   (Register: Short Story, CORE)
V.   Everyone here understands *that*/zero it's not easy. (Register: Informational Description, CORE)

The methodological framework for this paper was developed around a comprehensive review of *that*/zero variation literature which identified thirteen different structural features that have been claimed to predict the presence of the zero complementizer. This same process was used to identify the role(s) that different types of register also apparently play in complementizer selection processes. This review resulted in the formulation of the following argument/claim: the more formal the register (and genre) the more likely the presence/use of the overt *that* complementizer form.   To be more specific, authors have argued for a 'register continuum' whereby (and conversely) the greater the informality of both written and spoken texts (Bryant, 1962: Rissanen, 1991), the greater the degree of subjectivity in a text (Storms, 1966), the presence of a more casual style (Elness, 1984), the more informal the text type (spoken versus written, private versus public , unscripted/scripted) and the greater the informality of writing style (i.e. non-academic/academic and non-professional/professional) (Kaltenbock, 2004) are all claimed to act as predictive factors for the use/presence of the more 'informal' zero complementizer form. These structural and register based variables were then used to guide the study's methodological design, data coding parameters and research questions.

Using the 50 million word Corpus of Online Registers of English (CORE) a total of 45,000 hits (for all 5 verbs) were randomly extracted from each of CORE's 32 different register categories (Biber, Egbert & Davies, 2015). Once the data was extracted all of matrix plus complement that/zero constructions were coded for register category plus 28 structural variables including person, tense, polarity, and presence of modal auxiliaries, syntactic complexity, and complement clause subjects. Statistically sufficient sample sizes (n>40) for all register categories were extracted and a stepwise regression analysis is used to examine the statistical significance of 13 structural factors (as summarized in Kaltenbock 2004, described in Torres Cacoullos and Walker 2009 and presented in Author 1, Author 2 and Cuyckens, H, 2014) and register in regards to the selection of that/zero usage patterns for all five verbs in each of the 32 register categories.

The research questions guiding this paper are the following: (i) whether the conditioning factors proposed in the literature indeed predict the zero form, (ii) to what extent these factors interact, (iii) whether the predictive power of the conditioning factors

becomes stronger or weaker depending upon the type of register vis-à-vis formality, (iv) the impact of register itself as a variable on predicting the presence zero form, (v) what is nature of the relationship (if any) between the each of the five MSPs, the register categories and *that*/zero variation in general – do any MSPs show a tendency to be utilized within particular registers in terms of either significant *that* or zero complementizer usage and (vi) what new or additional insights are gained with this approach compared to previous studies regarding the impact of register on *that*/zero variation patterns conducted with significantly smaller corpora containing substantially fewer register categories and non-multivariate modelling (Finegan & Biber, 1995; Kaltenbock 2004).

The results reveal varying degrees of significance for each of the 13 matrix and complement clause features, however; stronger significance and implications are revealed when additional variables (e.g. polarity, length of the subject, verb type, the effect of register(s) as a variable etc.) are incorporated via a 'weighted' variable analysis. These findings are used to identify the structural factors which are significant in predicting the presence of the zero complementizer form for each of the 5 verbs and within and across each of the 32 register categories. These findings will then be used to facilitate a discussion concerning the implications for using this type of statistically driven multifactorial approach.

References

Author 1, Author 2 & Author 3. (2014) A Diachronic Corpus based Multivariate Analysis of "I think that" versus "I think zero" Polysemy and Synonymy. Corpus Methods and Applications in Cognitive Linguistics. Glynn, D & Robinson, J. (Eds). John Benjamins Publishing Co.

Biber, Douglas. (1988). Variation across speech and writing. Cambridge, UK: Cambridge University Press.

Biber, Douglas. (1999). A register perspective on grammar and discourse: variability in the form and use of English complement clauses. Discourse Studies 1. 131-50.

Biber, Douglas and Conrad, Susan. (2009) Register, Genre, and Style. Cambridge: Cambridge University Press.

Biber, Douglas, Egbert, Jessie and Davies, Mark (2015) Exploring the composition of the Searchable web: a corpus-based taxonomy of web registers. Corpora 10:1, 11-45

Bryant, Margaret M. (1962). Current American usage. New York: Funk & Wagnalls.

Davies, Mark. (2016-) Corpus of Online Registers of English (CORE). Available online at http://corpus.byu.edu/core/.

Elsness, Johan. (1984). "That or zero? A look at the choice of object clause connective in a corpus of American English", English Studies 65: 519-533.

Egbert, Jessie., Biber, Douglas. and Davies, Mark. (2015), Developing a bottom-up, user-based method of web register classification. Journal of the Association for Information Science and Technology, 66: 1817–1831.

Finegan, Edward; Biber, Douglas. 1995. "That and zero complementisers in Late Modern English: exploring ARCHER from 1650-1990", in Aarts, Bas; Meyer, Charles, M. (eds.). The verb in contemporary English. Theory and description. Cambridge: Cambridge University Press, 241-257.

Kaltenböck, Güner. 2004. "That or no that – that is the question. On subordinator suppression in extraposed subject clauses", Vienna English Working Papers 13 (1): 49-68.

Rissanen, Matti. 1991. On the history of that/zero as clause object links in English. In Aijmer, K. &B. Altenberg (eds.), English Corpus Linguistics; Studies in Honor of Jan Svartvik. London and New York: Longman. 272-289.

Storms, G. 1966. That-clauses in Modern English. English Studies 47. 249-70.

Torres Cacoullos, Rena & James A. Walker. 2009. On the persistence of grammar in discourse formulas: A variationist study of that. Linguistics 47. 1–43.