1st Call for H.F.R.I. Research Projects for the Support of Faculty Members and Researchers and the Procurement of high-cost Research Equipment

Deliverable 4: Lexicon and Related Algorithm for Content Analysis

Organizational Learning and Performance Outcomes: A Multilevel Longitudinal Study

ORGLEARN

HFRI-FM17-1799

Short description of the Deliverable 4

According to the project's latest amendment (Α.Π. 49347/18-05-2022), the Deliverable 4 refers to the development of a lexicon for content analysis to facilitate the secondary data collection.

Lexicon and Related Algorithm for Content Analysis

Our goal was to develop a lexicon to compute the relative amount of the variables of our conceptual model presented in Deliverable 2 in each company. To achieve this aim, we adopted the procedures -slightly altered to better serve our research design- of previous research relative to the development of lexicons utilized in content analyses (see, for instance, Uotila, Maula, Keil, & Zahra, 2009). The final measures should be available for most of the companies included in our dataset as well as applicable in most industries.

Traditionally, the coding of variables in text analysis was performed manually by researchers who were previously trained in this task. However, in recent years, coding using a computer software with a parser and a related, well-defined lexicon has gained popular attention due to the fact that it is equally accurate to human-performed coding (King & Lowe, 2003).

Additionally, content analysis methodology (i.e., counting and scoring words and their frequencies in texts) has been shown to provide valid and robust results reproducing those in more controlled environments such as laboratories (Porac, Wade, & Pollock, 1999; Uotila et al., 2009) as well as to measure concepts otherwise impossible to be captured (Tetlock, Saar-Tsechansky, & Macskassy, 2008).

The most important step in content analytic methodology is to create a lexicon with high levels of accuracy to the concepts of interest. To do so, we first derived the original definitions of our variables of interest (see Table 1; i.e., those variables that make sense to be measured through annual reports) and, then, created the word roots of these variables (see Table 2).

Original Definitions
Radical innovations refer to a fundamental departure from existing knowledge sources (Benner & Tushman, 2003; Levinthal & March, 1993), and are targeted at satisfying emerging needs or new markets by offering novel designs and breakthrough products and services (Jansen, Van Den Bosch, & Volberda, 2006).
Incremental innovations build on current knowledge and expertise (Benner & Tushman, 2003; Levinthal & March, 1993) and are targeted towards existing markets by improving current products and services (Jansen et al., 2006).

Potential Absorptive Capacity (PACAP) denotes knowledge acquisition (i.e., a team's ability to acquire externally generated knowledge that is critical to its operations) and knowledge assimilation processes (i.e., a team's routines that allow it to analyze, interpret, and understand the information obtained from external sources) (Zahra & George, 2002, p. 189).

Realized Absorptive Capacity (RACAP) captures knowledge transformation (i.e., a team's routines that facilitate combining existing knowledge and the newly acquired and assimilated knowledge) and knowledge exploitation processes (i.e., a team's routines that allow it to refine and extend existing competences or create new ones by incorporating acquired and transformed knowledge into its operations) (Zahra & George, 2002, p. 190).

Task conflict refers to differences in viewpoints and opinions among team members pertaining to group task (Jehn, 1995).

Knowledge breath refers to distinct and multiple sources of knowledge within teams (De Luca & Atuahene-Gima, 2007).

Table 1: Definitions of Variables of Interest

Variable of Interest	Words & Word Roots
*The star sign ' * ' can represent any characters.*	
Radical innovation	Radic*, novel*, breakthr*, discover*, search*, variation*, risk*, experiment*, play*, flexib*,
Incremental innovation	Increment*, improv*, refine*, choice*, select*, efficien*,
PACAP (i.e., knowledge acquisition and assimilation)	Acquisit*, collect* assimilate*, access*, gain*, analyz*, interpret*,
RACAP (i.e., knowledge transformation and exploitation)	Tranformat*, exploit*, implement*, execut*, use*, utiliz*, apply*
Task conflict	Conflict*, disagree*, disput*, quarrel*
Knowledge breadth	Breadth*, broad*, wide*, diverse*

Table 2: Words & Word Roots of The Variables of Interest

In a next step and to validate the accuracy of the newly developed instrument, we also assigned two PhD students to manually code the same variables in two distinct sets of company annual reports. The annual reports in these samples were derived from company websites as well as from the Compustat database. The sample of this analysis included 20 companies (randomly chosen) for the period 2010-2015, hence, comprising of 100 annual reports in total. The

interrater reliability of the classification process with two coders was kappa = 0.74 (Cohen, 1960). The frequency of each variable per thousand words in the annual reports is presented in Table 3 (Uotila et al., 2009).

At the same time, we also run an analysis using the computer-aided software and the list of word roots presented in Table 2. The correlation of the results of the manual coding and the automatic software in the same sample (i.e., of those 20 companies/ 100 annual reports) was high for most variables (see Table 3), thus, verifying the correct use of the automated method.

Variable of Interest	Correlation of Automated and Manual Measurement	Frequency (per 1000 words) in Manual Measurement
Radical innovation	0.59 ($p<0.05$)	5.5
Incremental innovation	0.67 ($p<0.01$)	2.3
PACAP (i.e., knowledge acquisition and assimilation)	0.51 ($p<0.05$)	3.1
RACAP (i.e., knowledge transformation and exploitation)	0.52 ($p<0.05$)	2.8
Task conflict	0.61 ($p<0.05$)	1.5
Knowledge breadth	0.55 ($p<0.05$)	2.9

Table 3: *Words & Word Roots of The Variables of Interest*

Hence, the aim of this lexicon is to help us measure the constructs of our survey at the firm-level of analysis by utilizing secondary data derived from companies' annual reports.

References

Benner, M. J., & Tushman, M. L. (2003). Exploitation, exploration, and process management: The productivity dilemma revisited. *Academy of Management Review, 28*(2), 238-256.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*(1), 37-46.

De Luca, L., & Atuahene-Gima, K. (2007). Market knowledge dimensions and cross-functional collaboration: Examining the different routes to product innovation performance. *Journal of Marketing, 71*(1), 95-112.

Jansen, J. J. P., Van Den Bosch, F. A. J., & Volberda, H. W. (2006). Exploratory innovation, exploitative innovation, and performance: Effects of organizational antecedents and environmental moderators. *Management Science, 52*(11), 1661-1674.

Jehn, K. A. (1995). A multimethod examination of the benefits and detriments of intragroup conflict. *Administrative Science Quarterly, 40*, 256-282.

King, G., & Lowe, W. (2003). An automated information extraction tool for international conflict data with performance as good as human coders: A rare events evaluation design. *International Organization, 57*(3), 617-642.

Levinthal, D. A., & March, J. G. (1993). The myopia of learning. *Strategic Management Journal, 14*(S2), 95-112.

Porac, J. F., Wade, J. B., & Pollock, T. G. (1999). Industry categories and the politics of the comparable firm in CEO compensation. *Administrative Science Quarterly, 44*(1), 112-144.

Tetlock, P. C., Saar-Tsechansky, M., & Macskassy, S. (2008). More than words: Quantifying language to measure firms' fundamentals. *The journal of finance, 63*(3), 1437-1467.

Uotila, J., Maula, M., Keil, T., & Zahra, S. A. (2009). Exploration, exploitation, and financial performance: analysis of S&P 500 corporations. *Strategic Management Journal, 30*(2), 221- 231.

Zahra, S. A., & George, G. (2002). Absorptive capacity: A review, reconceptualization, and extension. *Academy of Management Review, 27*(2), 185-203.