

Using Natural Language Processing in Entrepreneurship Research: Exploring the Perceived Benefits of Business Incubation

Benoit Gailly¹, Mahamadou Biga-Diambeidou^{2,1}, Christian Gnekpe¹, Thomas François^{3,4}, Patrick Watrin³ and Françoise de Viron¹

¹CRECIS, Université catholique de Louvain

²ICN Business School

³Centre for Natural Language Processing (Cental), Université catholique de Louvain

⁴Chargé de recherche FNRS

Policy makers have for many years taken for granted that implementing dedicated structures aimed at supporting start-ups had a positive impact on the development of those start-ups. Indeed, building on insights provided by the literature on National/Regional Innovation Systems (Lundvall 1992; Nelson 1993; Cooke et al. 1998; Asheim et al. 2003) and Incubations mechanisms (Etzkowitz and Leydesdorff, 2000; Hackett and Dilts, 2004; Phan et al., 2005; Bergek and Norman, 2008; Bruneel et al, 2012) policy makers perceive such “business incubators” as a way to promote regional economic growth through the support of young and innovative firms. The underlying assumption is that incubators help young firms to develop their innovative products, commercialize them, and maintain a sustainable growth during the start-up period.

As a consequence local policy makers have been and still are very enthusiastic in establishing and/or granting financial supports to incubators. For example, the National Business Incubation Association estimates that the number of incubators in the United States has increased from 12 in 1980 to 1,250 in 2012 and that there were about 7,000 such structures across the world.

However, there is surprisingly little empirical validation of the positive impact of those structures, with few or no evaluation of their long term performance in helping incubated firms (Vanderstraeten et al., 2014; Schwartz, 2011). As a result, there is evidence of potential

political pressure to keep some results under the radar (Fromhold-Eisebith and Eisebith, 2008; Amezcua, 2010; Bergek and Normann, 2008; Mian, 1997). Indeed, some studies indicate that among the new firms incubated, few became fast-growing firms and/or reached the status of middle-sized companies (Schwartz, 2011; Amezcua, 2010; Barringer, Jones and Neubaum, 2005; Phan, Siegel, and Wright, 2005). Besides the above-mentioned mixed research results (Amezcua, 2010; Phan, Siegel, and Wright, 2005), related methodological, theoretical, and empirical limitations (Yu and Nijkamp, 2009) raise criticism about incubator performance studies.

Furthermore, those studies rely upon either quantitative (accounting) data available about the incubated start-ups to measure their development, or the qualitative coding and interpretation of interviews using standard content analysis techniques (Lincoln and Guba, 1985). (See for example Barringer, Jones and Neubaum, 2005, with both suffering from the typical limitations of such approaches).

In particular it remains largely unknown whether incubation really helps overcoming the resource deficiencies that young firms face in the first years after their market entry (Hannan and Freeman 1984; Stinchcomb, 1965). Overall, an assessment of public policies regarding incubators is still needed (Bergek and Norrman, 2008; OECD, 2006; Mian, 1996a; Mian, 1997). Moreover, there are few studies investigating the impact of incubation on incubated ventures' performance from their own point of view (Fromhold-Eisebith and Eisebith, 2008; Schwartz, 2013; Vanderstraeten et al., 2014).

As a consequence, research call for more qualitative bottom-up approach (see Diez, 2001) to explore the impact of incubators' support on incubated firms. In particular, how do incubators create a context that helps firms bring together the various resources they need (Hellmann & Puri, 2002)?

In that context, this paper aims at contributing to a better understanding of the impact of incubators on incubated firms, analyzing the narrative based perception of those firms regarding the benefits they anticipated and experienced from that incubation. Furthermore, in order to minimize the bias typically encountered when using such approach (Neergaard & Ulhoi, 2006), we mobilize an original text analysis methodology that addresses some of the weaknesses identified in recent entrepreneurship research on the assessment of incubation mechanisms (Vanderstraeten et al., 2014; Slotte-Kock and Coviello, 2009; Hoang and Antoncic, 2003).

Method

There are strong technical obstacles to the evaluation incubators' performance, namely (1) the difficulty to disentangle the impact of a specific policy from other economic factors as well as the difficulty to assess "soft" goals such as the development of intangible capital (Vanderstraeten et al., 2014; Van Haeperen et al. 2009; Venkataraman, 2004), (2) the non-consensus on the criteria that make it possible to judge the success of incubation: number of jobs created, survival rate, access to social capital, development of innovations, etc. (Phan *et al.*, 2005), and (3) the lack of empirically validated measurement scales (Vanderstraeten and Matthyssens, 2010).

As a consequence, we chose to adopt an inductive bottom-up approach based on semi-directive interviews of an international sample of founders and managers of firms hosted by incubators. We conducted one to two hours interviews with entrepreneurs, founders or executive members of 20 startups incubated in 3 different incubators in Belgium (CEI), Luxembourg (Technoport) and France (Promotech). Those interviews were conducted in French, registered and transcribed by one our research team.

In contrast with the qualitative approach used by most interview-based management research, we used natural language processing (NLP) approaches (Fig.1), which aim at systematically extracting the main topics covered in our data without relying exclusively on the subjective interpretation of researchers. The NLP techniques mobilized are applied in two steps. First, automatic term recognition (ATR) was first used to identify, in our interviews, the domain-specific terms, that are assumed to be the most relevant ones for a topic-based approach. In a second step, these terms were projected into a large semantic space and automatically grouped into clusters of related words.

ATR is generally based on statistical information (Chung, 2003; Scott, 2008; Kit and Liu, 2008), on linguistic information (Bourigault, 1992; Dagan and Church, 1994), or on both (Drouin, 2003; Loginova et al., 2012). In this study, we selected the Scott (2008)'s method, which compares the raw frequency of a word in a specialized corpus (here, our interviews) with its frequency in a corpus representative of the common general language (here, a corpus of various types of texts, including literary works, newspapers articles, etc. described in François (2014)). The comparison is performed using a statistical test, in this case the chi-square test. We applied this method at two levels: simple words (i.e. string of characters separated by blanks) and multi-words. The detection of MWEs remains a challenging task in NLP and resorts on statistical association measures between adjacent words to decide whether they co-occur more often than expected by chance, which correspond to arbitrary word usages (Smadja, 1993, 143). We used the hybrid MWE extractor of Watrin and François (2011), which relies both on linguistic and statistical information.

In the second step, we trained a statistical language model based on the currently popular model of word embedding (Bengio et al., 2003), of which we used a variant, better fit for documents: the Paragraph vector (Le and Mikolov, 2014). It consists in a neural network

mapping words to high-dimensional vectors that usually have about 100 to 500 dimensions (300 in our case). This produced a semantic representation of the language in which the words or documents can be projected. In order to automatically detect the main topics in the interviews, we first trained such semantic space based on more than 20,000 articles from finance French newspapers (Boursier, etc.). In a second step, we concatenated the interviews from the same incubator and projected them into our semantic space. However, the dimensions of the space are not easily interpretable, so we instead compared the semantic relatedness of the interviews to each chapter (Fig.2) of a book about entrepreneurship (Fayolle, 2012), each chapter being representative of a given dimension of the entrepreneurship process. The similarity between the vectors representing a chapter and an incubator was computed with the cosine similarity. It should be also mentioned that our semantic space was created using only the words detected as relevant at the end of the step one of our methodology.

Preliminary Results and Implications

Preliminary analysis of our model looks promising. We obtain two lists, one for words and one for multi-word expressions (MWE), which include words from our interviews that are representative of the problematic of incubation. Through the ongoing research work, the clusters that emerged indicate interesting patterns of attributes related with the supports offered by incubators.

Those results (Fig. 3) show that this systematic approach can be used to interpret text-based data and highlight key patterns within that data. Moreover, it also allows quantitative subsample comparisons. As an illustration (Fig. 4) we have tested whether the replies from one incubator were significantly ($p=0.05$) different from the others. In this case we see for example that for the Technoport incubator, the theme 2 "L'entrepreneur" is significantly less

present in the corresponding interviews, while the theme 5 "social entrepreneurship" is significantly more present. This opens interesting avenue for further research regarding the differences between the perceived values added of different incubators, for example based on their governance, strategy and mission or business model, in line with the heterogeneity identified in the literature (Bruneel, Ratinho, Clarysse and, Groen, 2012; Mian, Fayolle & Wadid, 2012; Schwartz, M. and Hornych, C., 2010)

By achieving the aim of this study, we will provide important empirical verification and extension of the framework in the process of incubator evaluation by considering the importance of both tenants (Bruneel et al., 2012; Chan and Lau, 2005; Jungman et al., 2004) as part of the key incubator stakeholders that should not be ignored in measures of incubator evaluation (Vanderstraeten et al., 2014). Our work will thus contribute to the ongoing debate regarding the limitation related with the common practice of using one stakeholder perspective (for example, Abduh et al., 2007) insufficient to capture the performance of business incubators. We will also contribute on how to systematically provide with a framework that can help to enhance the tenant's perception of the impact of incubators.

The main contribution of this paper lies at the methodological standpoint. This paper provides a workable framework – in terms of structures and relations - that may help incubators in their task of assisting/training their tenants in developing the necessary skills required for the professionalization of new firms.

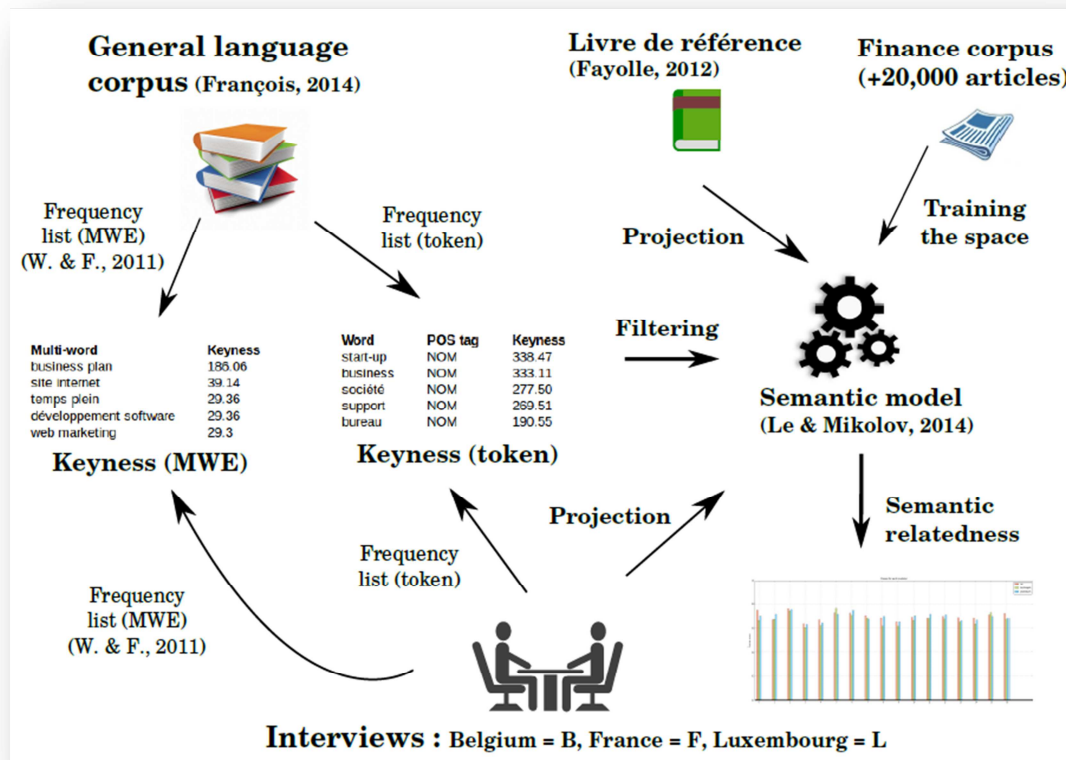


Fig. 1: Key steps of the natural language processing (NLP) method adopted in this study.

<u>List of chapters (labeled theme) from Fayolle, 2012</u>	
[0]	INTRODUCTION
[1]	L'ENTREPRENEURIAT, UN PHENOMENE ECONOMIQUE ET SOCIAL.
[2]	L'ENTREPRENEURIAT : CONCEPTION ET CADRE THEORIQUE ¹³
[3]	L'ENTREPRENEUR
[4]	CREATION DE VALEUR NOUVELLE ET INNOVATION
[5]	ENTREPRENEURIAT SOCIAL
[6]	CREER UNE ENTREPRISE OU UNE ACTIVITE
[7]	LA TRANSMISSION – REPRISE D'UNE ENTREPRISE OU D'UNE ACTIVITE
[8]	INTRAPRENDRE OU ENTREPRENDRE DANS UNE ORGANISATION
[9]	ENTREPRENDRE EN ÉQUIPE, BÉNÉFICES, PROCESSUS ET CONDITIONS DE SUCCÈS.
[10]	L'APPROCHE DU MARCHÉ ET LA DIMENSION MARKETING
[11]	L'APPROCHE ÉCONOMIQUE ET FINANCIÈRE DES PROJETS D'ENTREPRENDRE
[12]	L'APPROCHE JURIDIQUE ET CONTRACTUELLE DES PROJETS D'ENTREPRENDRE
[13]	LE BUSINESS PLAN : PROCESSUS D'APPRENTISSAGE DE LA STRATÉGIE ET CONVENTION SOCIALE
[14]	LE LANCEMENT DES ACTIVITÉS : L'ÉPREUVE DU FEU
[15]	DE L'ENTREPRENEUR AU MANAGER ENTREPRENEUR
[16]	LA CAPACITÉ STRATÉGIQUE DE L'ENTREPRENEUR ET LES STRATÉGIES DE DÉVELOPPEMENT DE LA JEUNE ENTREPRISE

Fig. 2: Description of the 16 themes (the chapters from Fayolle (2012) used as representative of each given dimension of the entrepreneurship process.

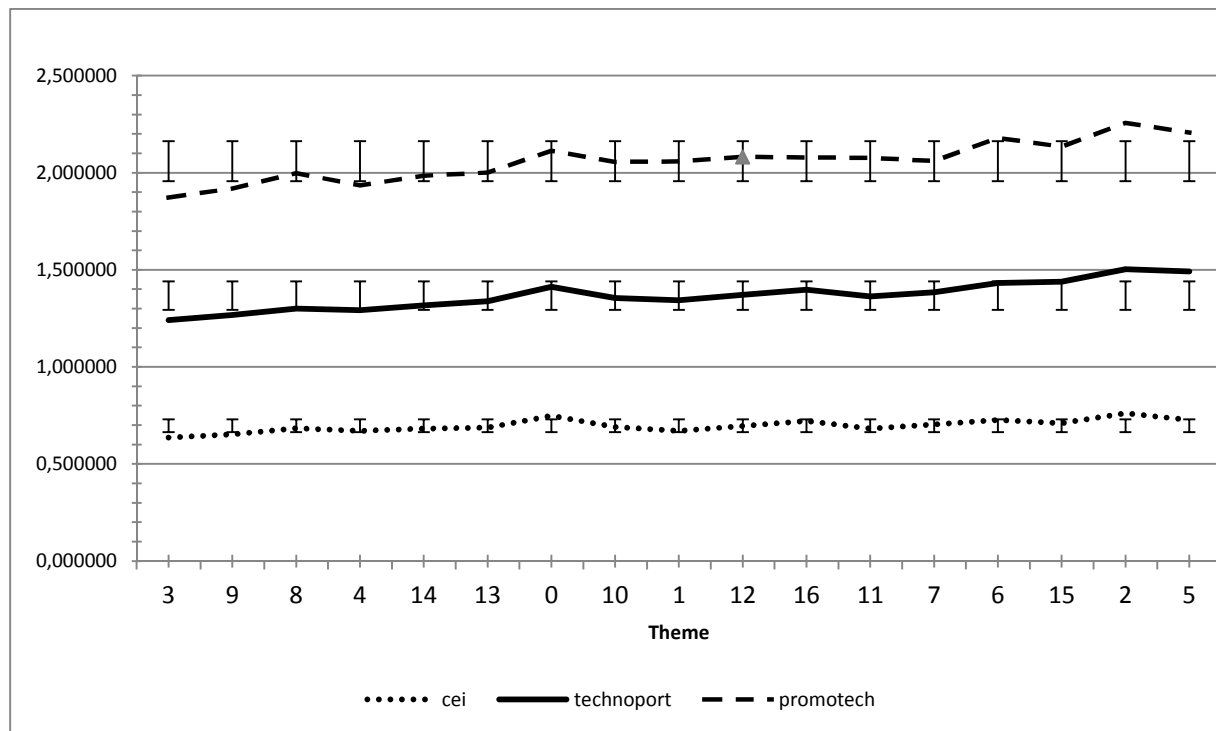


Fig 3: Distribution of patterns trough incubators

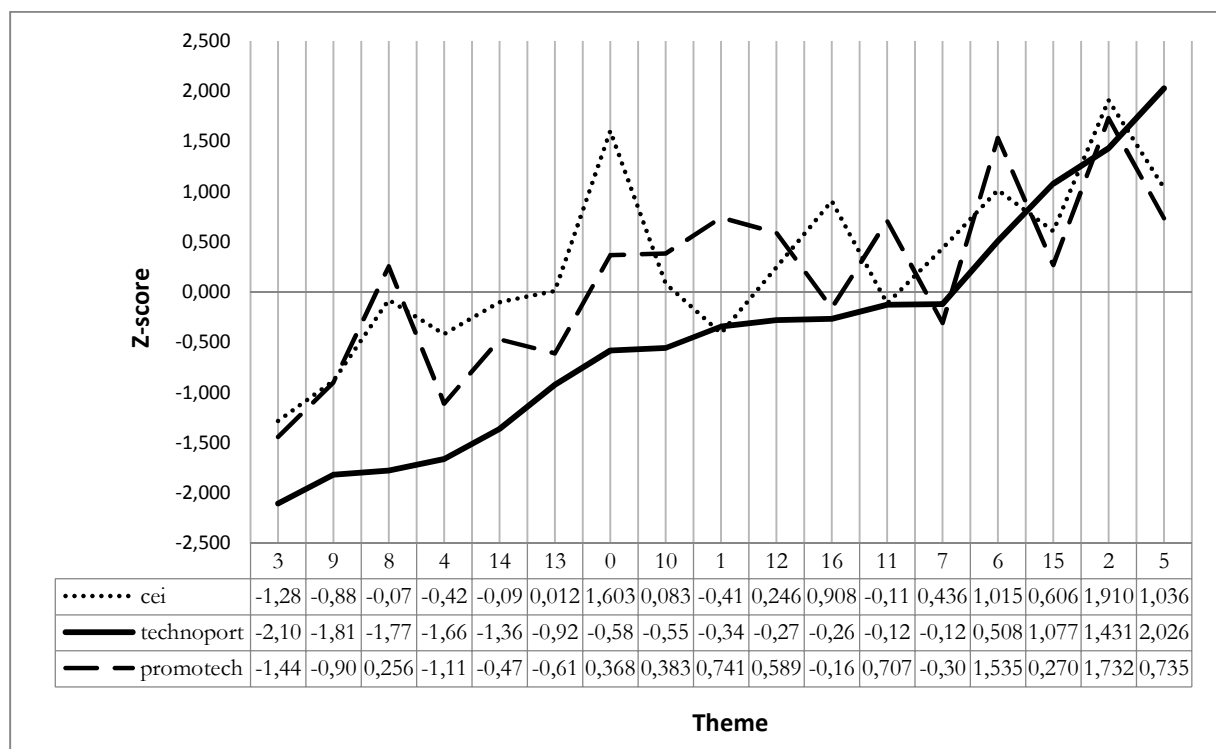


Fig 4: Key patterns of attributes related with the supports offered by incubators identified.

References

- Abduh, M., C. D'Souza., A. Quazi., H.T. Burley (2007). 'Investigating and Classifying Clients' Satisfaction with Business Incubator Services'. *Managing Service Quality*, 17, 74–91. doi:10.1108/09604520710720683
- Amezcu, Alejandro S. (2010) 'Performance Analysis of Entrepreneurship Policy: Which Business Incubators Generate the Highest Levels of Economic Performance?,' *Frontiers of Entrepreneurship Research*: Vol. 30: Iss. 18, Article 1. Available at: <http://digitalknowledge.babson.edu/fer/vol30/iss18/1>
- Asheim, B., A. Isaksen., C. Nauwelaers., F. Tötödling (2003). 'Regional Innovation Policy for Small Medium Enterprises'. Cheltenham, UK and Lyme, US : Edward Elgar
- Barringer, B. R., F.F. Jones, D.O. Neubaum (2005). 'A Quantitative Content Analysis of the Characteristics of Rapid-Growth Firms and their Founders'. *Journal of Business Venturing* 20(5), 663–687.
- Bengio, Y., H. Schwenk., J.S.Sen  cal., F. Morin., J. L. Gauvain (2006). 'Neural Probabilistic Language Models'. In *Innovations in Machine Learning* (pp. 137-186). Springer Berlin Heidelberg.
- Bergek, A., C. Norrman (2008) 'Incubator Best Practice: A Framework'. *Technovation*, (28), 1-2, 20-28. <http://dx.doi.org/10.1016/j.technovation.2007.07.008>
- Bourigault, D. (1992). 'Surface Grammatical Analysis for the Extraction of Terminological Noun Phrases' in proceedings of the Fifth International Conference on Computational Linguistics, pp. 977–81. Nantes, France.
- Bruneel J., T. Ratinho., B. Clarysse., A. Groen (2012). 'The Evolution of Business Incubators: Comparing Demand and Supply of Business Incubation Services across Different Incubator Generations'. *Technovation*, Vol: 32, Pages: 110-121, ISSN: 0166-4972.
- Chan, K. F., T. Lau (2005). 'Assessing Technology Incubator Programs in the Science Park: the Good, the Bad and the Ugly'. *Technovation*, Vol: 25, No: 10: 1215-1228.
- Chung, T.M. (2003). 'Identifying Technical Vocabulary', Unpublished PhD thesis. Wellington: Victoria University of Wellington.
- Cooke, P. (2005). 'Regional Asymmetric Knowledge Capabilities and Open Innovation Exploring 'Globalisation': a new model of industry organization'. *Research Policy*, Vol. 34, pp.1128–1149.
- Dagan, I., K. Church (1994). 'TERMIGHT: Identifying and Translating Technical Terminology' in proceedings of the Fourth Conference on Applied Natural Language Processing (ANLC 1994): 34-40.
- Diez, M. (2001). 'The Evaluation of Regional Innovation and Cluster Policies: Towards a Participatory Approach'. *European Planning Studies*, 9(7), 907-923.
- Drouin, P. (2003). 'Term Extraction using non-Technical Corpora as a Point of Leverage', *Terminology* 9 (1), pp. 99–117.

- Etzkowitz, H., L. Leydesdorff (2000). 'The Dynamics of Innovation: From National Systems and 'Mode 2' to a Triple Helix of University-Industry-Government Relation', *Research Policy*, Vol. 29, No. 2, pp.109–123.
- Fayolle A. (2012). 'Entrepreneuriat : Apprendre à Entreprendre', (2ème Ed), Dunod, pp358.
- François, T. (2014). 'An Analysis of a French as a Foreign Language Corpus for Readability Assessment', In Proceedings of the 3rd workshop on NLP for Computer-assisted Language Learning, NEALT Proceedings Series 22 / Linköping Electronic Conference Proceedings 107: 13-32.
- Fromhold-Eisebith, M., G. Eisebith (2008). Looking Behind Facades: Evaluating Effects of (Automotive) Cluster Promotion'. *Regional Studies* 42 (10), 1343-1356.
- Hackett, S.M., D.M. Dilts (2008). 'Inside the Black Box of Business Incubation: Study B – Scale Assessment, Model Refinement, and Incubation Outcomes. *The Journal of Technology Transfer*, Vol. 33, No. 5, pp.439–471.
- Hannan, M.T., J. Freeman (1984). Structural Inertia and Organizational Change. *American Sociological Review*, 49, 149-164
- Hellmann T., M. Puri (2002). 'Venture Capital and the Professionalization of Start-Up Firms: Empirical Evidence'. *The Journal of Finance*, Vol. LVII, No. 1, FEB. 2002.
- Hoang H., B. Antoncic (2003). 'Network-Based Research in Entrepreneurship: A Critical Review.' *Journal of Business Venturing* 18, 165–187.
- Jungman, H., J. Okkonen., T. Rasila., M. Seppä (2004). 'Use of performance measurement in V2C activity', *Benchmarking: An International Journal*, Vol. 11, No. 2, pp.175–189.
- Kit, C., X. Liu (2008). 'Measuring Mono-Word Termhood by Rank Difference via Corpus Comparison'. *Terminology* 14 (2), pp. 204–29.
- Le, Q.V., T. Mikolov (2014). 'Distributed Representations of Sentences and Documents'. *arXiv preprint arXiv:1405.4053*.
- Lincoln, Y.S., E.G. Guba (1985). 'Naturalistic Inquiry', Sage, Beverly Hills, CA.
- Loginova, E., A. Gojun., E. Blancafort., M. Guegan., T. Gornostay., U. Heid (2012). 'Reference Lists for the Evaluation of Term Extraction Tools'. *TKE 2012: Terminology and Knowledge Engineering*. 19–22 June.
- Lundvall, B-Å. (ed.) (1992). 'National Systems of Innovation: Towards a Theory of Innovation and Interactive Learning'. London: Pinter Publishers.
- Mian, S., A. Fayolle., L. Wadid (2012). 'Building Sustainable Regional Innovation Platforms for Incubating Science & Technology Businesses: Evidence from the US and French Science & Technology Parks'. *International Journal of Entrepreneurship and Innovation*, Vol 13, No. 4, pp 235-248
- Mian, Sarfraz (1997). 'Assessing and Managing the University Technology Business Incubator: an Integrative Framework'. *Journal of Business Venturing*, 12(4) 251-340.

- Mian, Sarfraz (1996). Assessing Value-Added Contributions of Technology Business Incubators to their Tenants. *Research Policy*, 25, 325-335.
- Neergaard H., J.P. Ulhoi (2006). 'Government Agency and Trust in the Formation and Transformation of Interorganizational Entrepreneurial Networks'. *Entrepreneurship Theory and Practice*, 30 (4), 519-539.
- Nelson R.R. (1993). 'National Innovation Systems: A Comparative Analysis'. Oxford University Press, pp560.
- Phan, P.H., D.S. Siegel., M. Wright (2005). 'Science Parks and Incubators: Observations, Synthesis and Future Research'. *Journal of Business Venturing*, Vol. 20, No. 2, pp.165–182.
- Schwartz, M. (2013). 'A Control Group Study of Incubator's Impact to Promote Firm Survival'. *Journal of Technology Transfer*, Vol. 38, No. 3, pp.302–331.
- Schwartz, M., M. Göthner (2009a). 'A Multidimensional Evaluation of the Effectiveness of Business Incubators: an Application of the PROMETHEE outranking method'. *Environment and Planning C: Government and Policy*, Vol. 27, No. 6, pp.1072–1087.
- Schwartz, M., M. Göthner (2009b). 'A Novel Approach to Incubator Evaluations: The PROMETHEE Outranking Procedures. IWH-Discussion Papers, January, No. 1.
- Schwartz, M., C. Hornych (2010). 'Cooperation Patterns of Incubator Firms and the Impact of Incubator Specialization: Empirical Evidence from Germany'. *Technovation*, Vol. 30, No. 9, pp.485–495.
- Scott, M. (2008). 'WordSmith Tools version 5'. Liverpool: Lexical Analysis Software.
- Slotte-Kock S., N. Coviello (2009). 'Entrepreneurship Research on Network Processes: A Review and Ways Forward'. *Entrepreneurship Theory and Practice*. Volume 34, Issue 1, pages 31–57, January 2010.
- Smadja F. (1993). 'Retrieving collocations from text: Xtract'. *Computational Linguistics*, 19:143–177.
- Stinchcombe Al (1965). 'Organizations and Social Structure'. In: March J editor editors. *Handbook of Organizations*. Chicago: Rand McNally; 1965. p. 153–193.
- Vanderstraeten J., P. Matthyssens (2010). 'Measuring the Performance of Business Incubators: A Critical Analysis of Effectiveness Approaches and Performance Measurement Systems'. ICSB World Conference Proceedings.
- Vanderstraeten J., P Matthyssens., A van Witteloostuijn (2014). 'Toward a Balanced Framework to Evaluate and Improve the Internal Functioning of Non-Profit Economic Development Business Incubators. A study in Belgium.'. *International Journal of Entrepreneurship and Small Business*, 23 (4), 478-508.
- Venkataraman S. (2004). 'Regional Transformation Through Technological Entrepreneurship'. *Journal of Business Venturing*, 19 (2004) 153–167.

- Watrin, P., T. François (2011). 'An N-gram Frequency Database Reference to Handle MWE Extraction in NLP Applications'. In Proceedings of the 2011 Workshop on MultiWord Expressions: from Parsing and Generation to the Real World (ACL Workshop), Portland, Oregon, June 23, 2011, 83-91.
- Yu, J. B., P. Nijkamp (2009). 'Methodological Challenges and Institutional Barriers in the Use of Experimental Method for the Evaluation of Business Incubators: Lessons from the US, EU and China. 2009 Atlanta Conference on Science and Innovation Policy, Atlanta, Georgia, IEEE.

Reproduced with permission of copyright owner. Further reproduction
prohibited without permission.