



A Search for Discriminative Linguistic Markers in ICT
Practitioner Discourse, for the Ex Ante Identification of
Disruptive Innovation

MMIM592

by

Ian Bloodworth
198509320

Supervisor : Jean-Gregoire Bernard

Submitted to the School of Information Management,
Victoria Business School,
Victoria University of Wellington
in partial fulfilment of the requirements for the degree of

Master of Information Management

October 2012

Preface

I would like to thank my wife Jacquie and beautiful daughters Isobel and Eliza for their patience, encouragement and support during the course of my studies, and in the writing and preparation of this paper. Without their support, this would not have been possible.

I would also like to thank my supervisor at Victoria University, Dr Jean-Gregoire Bernard, for his patience, advice, and invaluable input into this study.

Finally, I would like to dedicate this work to the memory of my father, who sadly passed away during the course of this work.

Abstract

Disruptive innovations have the potential to disrupt markets, and drive them in new directions. A common problem faced by business organizations is identifying such disruptive innovations. From a managerial perspective, there is real value in being able to accurately identify disruptive innovations early in the product life-cycle, as it affords the organization the opportunity to put in place business strategies that leverage this information, to gain maximal competitive advantage.

This investigation was undertaken to determine if linguistic markers could be identified in ICT practitioner discourse that could be used to discriminate between traditional business intelligence (BI) - the legacy or incumbent technology, and software-as-a-service (SaaS) BI - a new technology and candidate disruptive innovation.

Quantitative content analysis undertaken using the tool Veneficium WordFrequencyCounter, was used to analyze written practitioner discourse identified from within the Industry Newsgroup file of LexisNexis Academic universe. Analysis was undertaken using attribute sets derived deductively from the academic literature, and inductively from the data itself, which provided both manifest and latent meaning of component words. Individual relative word associations with both the traditional BI and SaaS BI corpora were also analyzed.

Analysis of the attribute set usage data provided evidence that manifest and latent word meaning remained consistent for the time period investigated in this study (2000 to 2012), and so could support the purpose of this study, and was suggestive of the fact that SaaS BI could be a disruptive technology. The study also identified that there was a significant difference in vendor and industry attribute set usage between the SaaS BI and traditional BI corpora, consistent with the Abernathy-Utterback model.

Analysis of individual word associations with the traditional BI and SaaS BI corpora identified a number of word association patterns that could discriminate between traditional BI and SaaS BI that may be transferable to other technologies. A crossover event pattern was also identified (in which the word association pattern switches between the incumbent and new technology), which may be able to provide an indication that a technology innovation is, or is about to become, disruptive.

This study contributes a new approach for investigating disruptive innovation, and highlights the potential of using content analysis of practitioner discourse to identify linguistic markers for disruptive innovation.

The key contribution of the study is the observation that discriminative linguistic markers can in fact be identified, and that such markers appear to have predictive capabilities. That is, they may allow organizations to identify disruptive innovations ex ante.

Table of Contents

1	Introduction	5
2	Literature Review	7
2.1	Technological Evolution and Innovation.....	7
2.1.1	Patterns of Technology Evolution.....	7
2.1.2	Henderson-Clark Model.....	10
2.1.3	S-curves.....	11
2.1.4	S-curves - background.....	12
2.1.5	Technology S-curves	13
2.1.6	The Teece Model (Dynamic Capabilities Framework)	16
2.1.7	Abernathy-Utterback Model.....	18
2.1.8	Abernathy-Utterback Model Extensions.....	21
2.1.9	Disruptive Innovation Theory	23
2.2	Software as a Service	33
2.3	Business Intelligence.....	33
2.4	SaaS BI.....	34
3	Methodology	37
3.1	Nature and Purpose of Content Analysis.....	37
3.2	Content Analysis Process	40
3.2.1	Data Collection.....	41
3.2.2	Coding & Data Presentation.....	46
3.2.3	Analysis and Interpretation.....	50
4	Results & Analysis	54
4.1	Article Volumes	54
4.2	Attribute Set Stability.....	55
4.3	Longitudinal Analysis of Attribute Sets.....	57
4.3.1	Vendor Attribute Set Differences	61
4.3.2	Industry Attribute Set Differences	62
4.4	Potential Linguistic Markers.....	64
4.4.1	Unique words.....	64
4.4.2	Words common across both corpora	67
5	Discussion	77
5.1	General Discussion.....	77
5.2	Limitations of Current Study.....	79
5.3	Future Research Opportunities.....	81
6	Conclusion	81
7	References	83

List of Figures

Figure 2.1. Innovation Continuum	9
Figure 2.2. Henderson-Clarke Framework for Defining Innovation.....	11
Figure 2.3. Sigmoid Function.....	12
Figure 2.4. Hypothetical S-curves.....	14
Figure 2.5. The Dynamic Capabilities Framework.....	18
Figure 2.6. Abernathy-Utterback Model.....	19
Figure 2.7. Hypothetical S-curve and Abernathy-Utterback Model.....	20
Figure 2.8. Disruptive Innovation.....	25
Figure 2.9. SaaS BI Conceptual Framework.....	35
Figure 2.10. Gartner Hype Cycle for Business Intelligence, 2011.....	36
Figure 3.1. Content Analysis Method Typology.....	40
Figure 3.2. Reference Text Categories for Content Analysis.....	44
Figure 4.1 Publication Volumes between 2000 & 2012.....	55
Figure 4.2. Cross sectional comparison of attribute sets composed of unique words.....	56
Figure 4.3. Cross sectional comparison of attribute sets composed of shared words.....	56
Figure 4.4. Attribute category abundance across the corpus for Traditional BI	57
Figure 4.5. Attribute category abundance across the corpus for SaaS BI.....	57
Figure 4.6. Relative abundance of attribute sets for traditional BI.....	58
Figure 4.7. Relative abundance of attribute sets for SaaS BI.....	58
Figure 4.8. Relative abundance of the Vendor attribute set.....	61
Figure 4.9. Distinct vendor numbers for SaaS BI and traditional BI.....	62
Figure 4.10. Relative abundance of the Industry attribute.....	63
Figure 4.11. Distinct industry numbers for SaaS BI and traditional BI.....	64
Figure 4.12. Relative frequency of attribute sets composed of unique words.....	65
Figure 4.13. Tag cloud for high frequency words shared between the corpora.....	70
Figure 4.14. Relative word association to both SaaS BI and traditional BI.....	71
Figure 4.15. Relative association of key words with SaaS BI and traditional BI over time.....	73

1 Introduction

A common problem faced by business organizations is identifying technological innovations that have the potential to disrupt the market and drive it in a new direction. Such technological innovations (referred to as disruptive innovations from this point forward) have managerial importance for both technology producing and technology consuming organizations, as they are a principal means of taking competitors by surprise and enhancing competitive advantage. The value of forecasting disruptive innovations lies in increasing the lead time for identifying such innovations; it promotes better decision making and provides additional stability to the organizational environment (National Research Council, 2010) .

This study investigated whether linguistic characteristics were present in Information and Communications Technology (ICT) practitioner discourse that could be used to identify the various categories of innovation. The motivation behind the investigation was to discriminate between disruptive innovation and non-disruptive innovation (e.g. incremental innovation). From a managerial perspective, the ability to identify disruptive innovation early in the product life cycle provides organizations the opportunity to adopt business strategies to leverage this information to gain competitive advantage. The research question addressed by this research is:

Can discriminative linguistic markers be identified in ICT practitioner discourse for the identification of disruptive technologies (ex ante), using SaaS BI as a test case?

Content analysis was used to explore the possibility of using practitioner dialogue published in the public domain (made available via LexisNexis) to identify generic discriminatory linguistic markers that could be used to distinguish between disruptive and non-disruptive innovations, with the aim of using these markers to determine whether Business Intelligence Software as a Service (SaaS BI) is a disruptive innovation.

SaaS BI has been described by Gartner as those BI functions and applications supported by a vendor as a service, accessible via the internet, and without the need to deploy or maintain an on-premises solution (Bitterer et al., 2007). SaaS BI was chosen as a case for this investigation, because it is still relatively early in its product life-cycle and there is practical value in identifying if it is a disruptive innovation.

This investigation begins with a review of the literature relevant to innovation, Software as a Service (SaaS), business intelligence, and SaaS business intelligence. Subsequent sections describe the

analytical methodology followed; results obtained; a discussion of findings; and possible avenues for future research. Finally, the conclusion summarizes the outcome of this research.

2 Literature Review

This review summarizes prior and current literature relevant to the topic of investigation. It provides a theoretical basis to aid understanding technological evolution and the concept of disruptive innovation, and its managerial implications. It then describes business intelligence, with a focus on SaaS BI. A description is then given regarding how assessments have been made in the academic literature to identify disruptive innovations. Finally, a high level review of content analysis is provided, including examples of prior social science research utilizing content analysis.

2.1 Technological Evolution and Innovation

The following sections provide an overview of technological evolution and innovation. The sections are organized in approximate chronological order of when the sections' key concepts appeared in the academic literature. There is also a distinction between static and dynamic models. All models discussed up to the Teece model (with the exception of S-curves) can also be classified as static models, models after this point can be considered dynamic (Chang, Lai, & Chang, 2009). The main difference between static and dynamic models is that dynamic models incorporate the concepts of a life-cycle, describing the dynamic process of technology creation through to technological obsolescence.

2.1.1 Patterns of Technology Evolution

The concept of technological evolution can be traced back to the Schumpeterian trilogy of invention (generation of new ideas), innovation (the further development of those ideas into marketable products and processes), and diffusion (the process whereby the product or process spreads throughout its potential market) which Schumpeter used to explain business cycles (Stoneman & Diederer, 1994). One of the key features of Schumpeter's idea is that technological evolution is a process, and not just an idea, although the implied linearity of the process ran counter to the later cybernetic viewpoint (Bunnell & Coe, 2001). This viewpoint emphasizes systems and communication loops, and argues that the evolutionary process is enacted through webs of social relations as opposed to isolated events associated with individual researchers or entrepreneurs. It also needs to be kept in mind that Schumpeter's focus was macroeconomic business cycles, not the microeconomic focus of individual organizations.

Business organizations display varying degrees of commitment to "search" for and develop innovations related to contemporaneous problems particular to a given set of researchers; that is,

they undertake specific, targeted problem solving activities. Such activities are based upon a mix of exogenous scientific advances and endogenously accumulated knowledge, tempered by the appropriability of the knowledge in the organizational context (Dosi, 1988). Typically, these searches are motivated by market driven profit seeking by the organization, and are most often shaped by a given technological paradigm; technological paradigms constrain the search activity, by contextually defining the techno-economic need to be fulfilled, the scientific principles utilized for the task -i.e. a "model", and a "pattern" of enquiry (Dosi, 1982). While technological paradigms are useful for focusing an organization's research and development efforts (Shane, 2008, p.19), they can also be limiting, in that they can blinker researchers, and prevent them from investigating alternative, possibly superior, solutions (Dosi, 1988).

Related to the concept of technological paradigm is the concept of technological trajectories. Such trajectories can be envisioned as iterations of the problem solving activity determined by a technological paradigm, and are what provide the momentum for the evolution of a particular technology (Nelson & Winter, 1977). An important feature of technology trajectories is that they possess inertia; that is, if a trajectory is powerful, it may be difficult for an organization to switch from one trajectory to another, particularly if the technological paradigms are different. And furthermore, when an organization changes trajectories, they are often required to start the problem solving activity from the beginning (Dosi, 1982). The relevance of this feature will be discussed later.

Nelson & Winter (1977) defined innovation as a "portmanteau" blending the wide range of variegated processes by which technology evolves over time. This definition is entirely consistent with an earlier definition of innovation provided by Zetman, Duncan and Holbek in 1973 (cited in *Dewar & Dutton, 1986*): "an idea, practice or material artifact perceived to be new by the relevant unit of adoption". Dewar & Dutton (1986) extended this definition to include the idea that innovations can vary in their degree of "newness" to the adopting unit, and used the notion of radicalness to capture this. Radical and incremental are used to describe different types of technological innovations; radical innovations involve revolutionary changes in technology and involve significant departure from existing practice, while incremental innovations represent minor improvements or adjustments to existing technology.

While it is difficult to accurately assess the true rate of radical innovation, most authors agree that almost all technological innovations are incremental (Shane, 2008, p.21), although the much lower frequency radical innovations have significantly greater impact, due to the fact that they are high-risk high-return developments. Also, the organizational requirements for effectively managing

incremental innovation are different from those required for managing radical innovation, and involve differing strategies and organizational structures (William, 2006).

Figure 2.1 illustrates the continuum of innovation, from low risk incremental innovation, through to high risk disruptive (breakthrough) innovation. The more radical an innovation, the more difficult it is to estimate its market potential and acceptance; consequently its associated market acceptance and financial risk is greater (Assink, 2006). Of interest to this work is disruptive and incremental innovation. Breakthrough innovations are those innovations that involve unexpected leaps of creativity and insight, generating novel products and creating new markets (Mascitelli, 2000). Often, breakthrough innovations are also referred to as radical innovations as described previously (Inauen & Schenker-Wicki, 2012). As will be discussed later, such radical (breakthrough) innovations are favored in environments with small entrepreneurial companies who are engaged in innovation competition, consistent with Schumpeter's view of innovation (Pyka & Andersen, 2012).

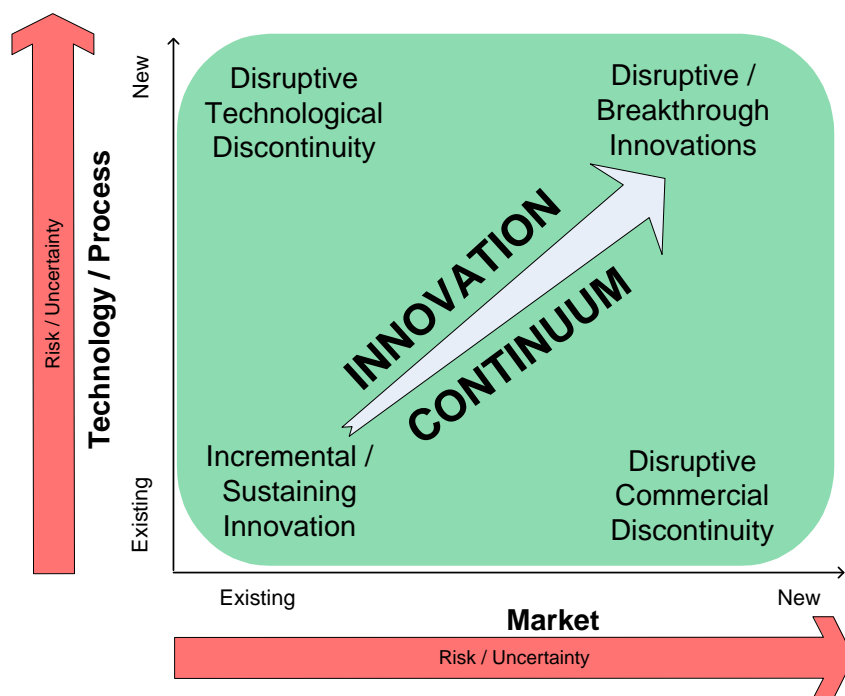


Figure 2.1 Innovation Continuum

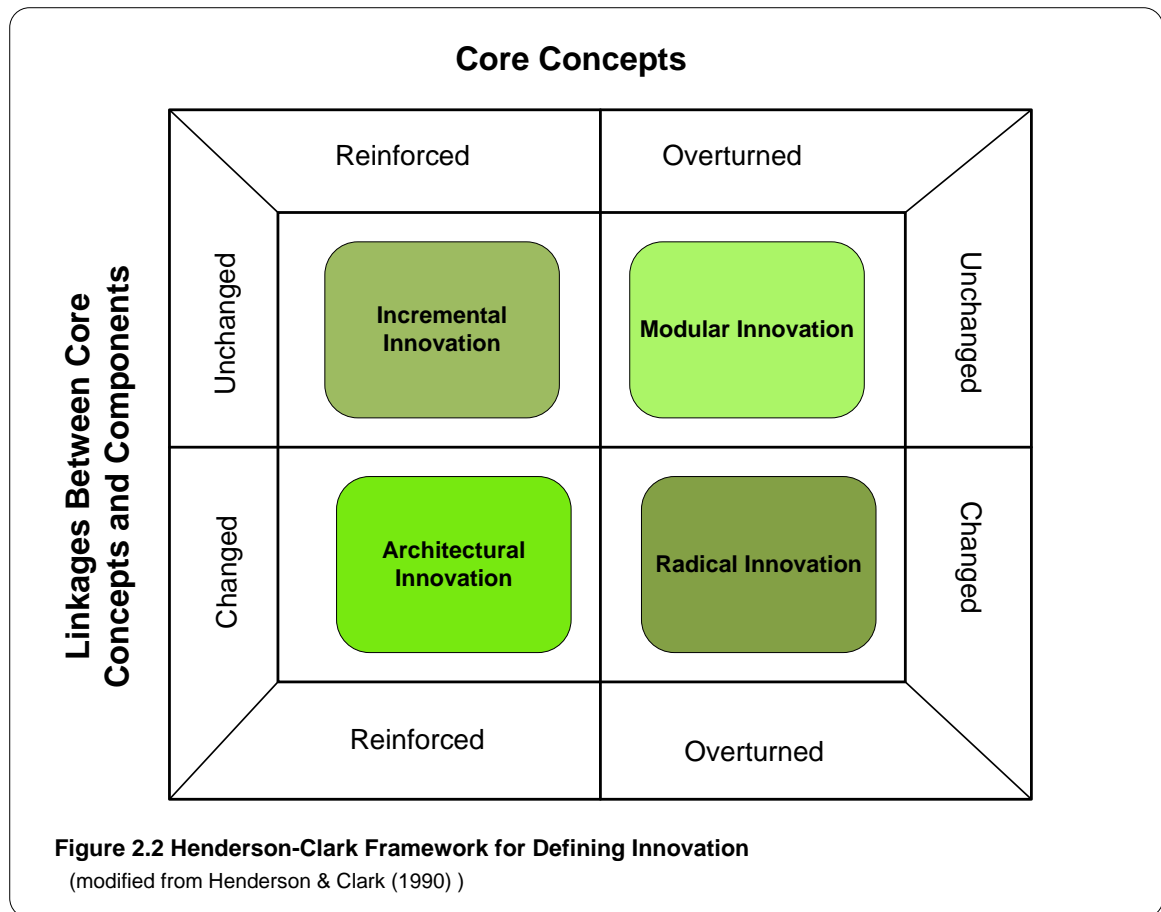
2.1.2 Henderson-Clark Model

The concept of radical and incremental innovation was further refined by Henderson & Clark (1990), who realized that on their own, these concepts could not predict organizations in the better position to innovate under a particular set of circumstances. Henderson and Clark added a further two dimensions to innovation; component knowledge and architectural knowledge. Component knowledge related to knowledge about physically distinct parts of a "product" that embodied a core design concept and performed a defined function, while architectural knowledge related to knowledge of the linkages that linked components together.

The Henderson-Clark model has four kinds of innovation, as shown in figure 2.2. In this model, radical innovation establishes a new dominant design, and thus a new set of core design concepts from which components are derived, and linked together in a new architecture. Incremental innovation extends the existing design, with improvements occurring in individual components, but the linkages between components remain the same. The two new types of innovation introduced by this model are modular innovation, whereby only the core design concepts of a technology are changed (this differs from incremental innovation by the fact that components are changed and not extended), and architectural innovation, where only the linkages between components are changed, without significant change to the existing components themselves.

A key implication of this model is that organizations must distinguish between incremental, modular, and architectural innovation, as the strategies (and indeed organizational competencies) to exploit one type of innovation may not be appropriate for the others.

Additionally, Henderson and Clark describe how architectural innovations (especially radical) can be challenging for any market's key incumbent 'suppliers', both in terms of instigating and recognizing. One of their reasons for this is that once a product innovation has reached the dominant design stage, and the market reaches the specific phase (terms taken from the Abernathy-Utterback model, discussed later), the product architecture can become embedded within the organization, and architectural component specific business units or teams can embed in to the organizational structure. From an organizational perspective, this approach promotes both incremental (sustaining) innovations and also the development of deeper domain knowledge; both these factors however, make it difficult for organizations to develop or respond to radical innovations, and contribute to the concept of technology trajectories possessing inertia. Although not specifically framed as such in their paper, the embedding of architectural component specific business units within the organizational structure also strengthens the organization's cultural inertia, which is equally difficult to overcome, and tends to inhibit disruptive innovations (Assink, 2006).



The models discussed up to this point have a limitation in that they cannot be used to predict when a technological discontinuity will occur. A useful technique to visualize technical innovation for any type of technology, and analyze technology lifecycles, is to plot a technology S-curve for the technology in question. Such curves are discussed in the following section.

2.1.3 S-curves

One of the first people to describe a mathematical variant of S-curves was Benjamin Gompertz, who in the first half of the 19th century developed a mathematical model for time series, where growth (of some measure) is slowest at the start and at the end of a period (Winsor, 1932). Soon after its development, this mathematical model was applied to a wide variety of fields, and today, is widely found in both natural and social science research.

For the current study, the relevance of S-curves is that they present a useful visual framework for illustrating technology cycles, and the difference between incremental and disruptive innovation. Identifying disruptive innovations is a key goal of this study.

This section provides background information on S-curves generally, and also on technology S-curves in particular.

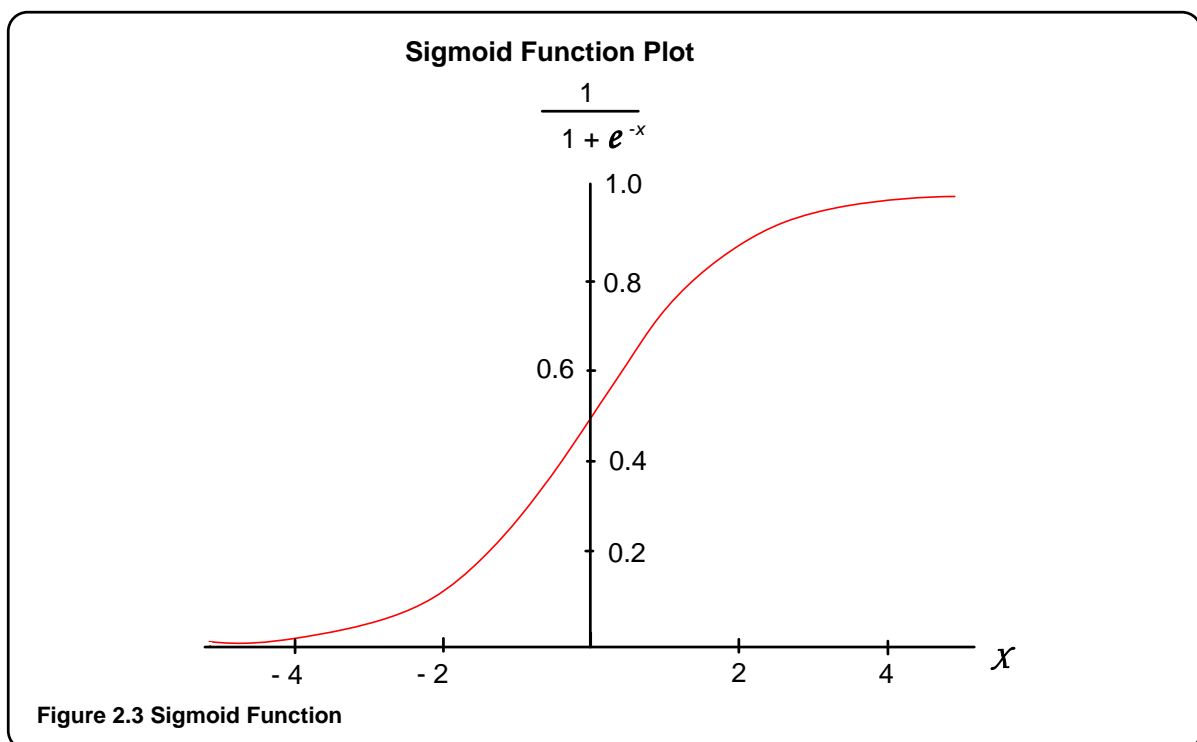
2.1.4 S-curves - background

S-curves are a variant of the sigmoid function, a mathematical function that produces a sigmoid curve. As suggested by their name, S-curves have a characteristic "S" shape.

Typically, sigmoid functions are logistic functions, but can also include arctangent, hyperbolic tangent, and generalized logistic functions (Huettenmueller, 2006, p.288). While a detailed mathematical description will add little value to this review, the following high level information (relating to a logistic function) is provided to aid later discussion.

The definition of the sigmoid function $[x]$ is: _____

In the expression above, e is Euler's number, and x is some variable such as time (or effort, cost, etc). The plot of this function (displaying the typical S-curve shape) is shown in figure 2.3.



As can be seen, after a period of slow growth, the function shows growth approximating exponential, then growth begins to slow, and ultimately the growth rate approaches zero. In general, the integral of any smooth, regular, positive "bell shaped" function will be sigmoidal; therefore the cumulative distribution function for normal distributions will result in an S-curve

(Bowling, Khasawneh, Kaewkuekool, & Cho, 2009). The significance of this point is shown in figure 2.4, and discussed later.

S-curves are common in natural sciences. For example, Cole & Green (1992) demonstrate that S-curves model triceps skin-folds and body weight in humans; in their paper Simons & Chow (2012) show very typical sigmoidal dose response curves for steroid hormone activity; Martin & Lubow (2011) demonstrate the use of an S-curve in characterizing phenomena observed in stellar accretion discs; and Kongparakul, Ng, & Rempel (2011) show that the catalyzed metathesis hydrogenation of deproteinized natural latex follows an S-curve.

S-curves also frequently appear in social science research. For example, Ransbotham & Mitra (2011) show that an S-curve can be used to model the diffusion of computer system attacks; in studying nonlinear probability weighting, Levine (2012) demonstrated that probability weighting curves in neural networks can follow an S-curve; Sima, Panageas, Heller, & Schrag (2010) demonstrate how the diffusion (market penetration) of chemotherapy drugs, post US FDA approval (and at the aggregate level) can be modeled using an S-curve; and Ghosh (2012) demonstrated that the relationship between U.S. trade balance and terms of trade could be modeled as an S-curve.

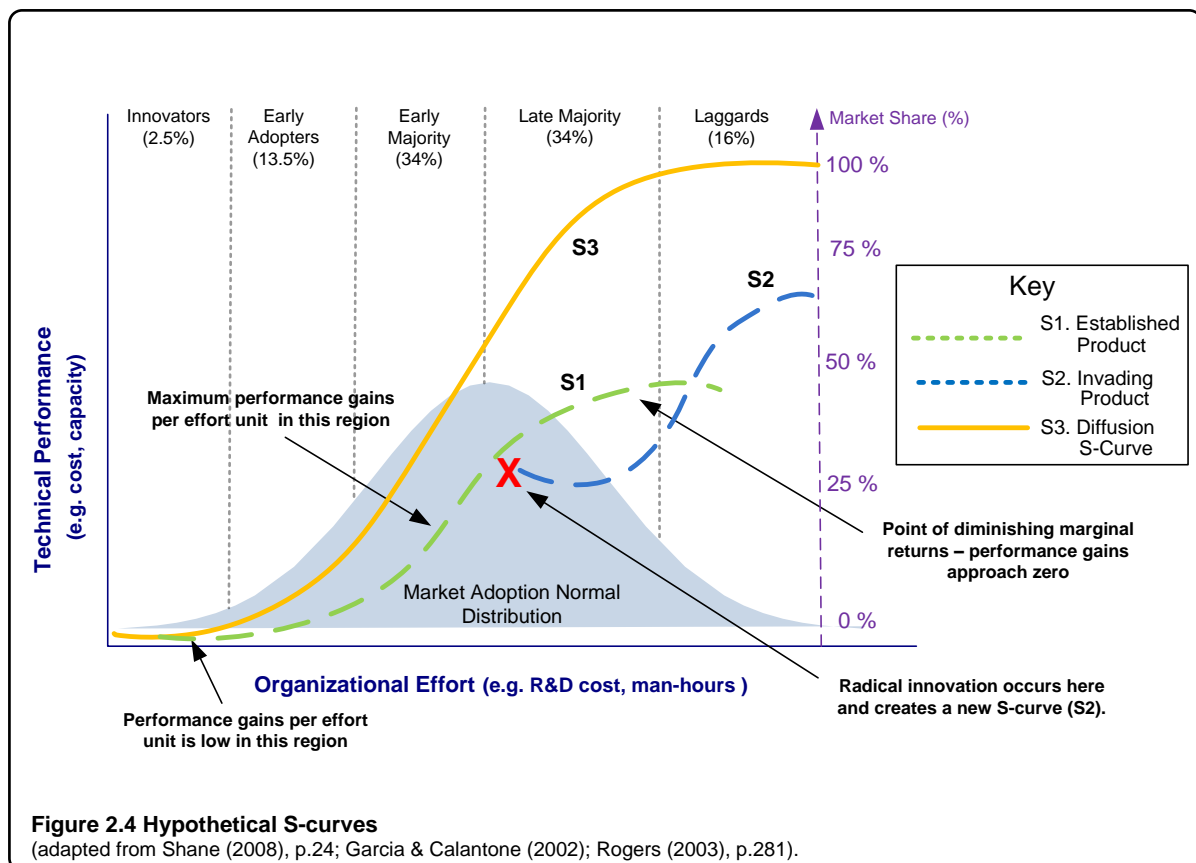
A possible explanation for the near universality of S-curve patterns in the natural and business worlds was provided by Adrian Bejan, of Duke University in 1996. Bejan put forward a theory (the Constructal Law) that attempted to explain the prevalence of the S-curve. The law states: " For a finite-size flow system to persist in time (to live), it must evolve such that it provides greater and greater access to the currents that flow through it." (Adrian Bejan & Lorente, 2010, p.1335). Bejan based his theory on natural flow structures, and the fact that they contain two flow mechanisms; convection and diffusion (Bejan & Lorente, 2011). Bejan describes how flow systems spread by smaller branches bifurcating out of larger branches, essentially reducing and redistributing any forms of resistance; the larger branches deal with convection, the smaller branches diffusion. In a recent paper, Bejan uses the concept of flow systems to model the citation history in academic literature (Bejan & Lorente, 2012); this concept would apply equally well to technology diffusion.

2.1.5 Technology S-curves

A technique useful for visualizing the evolution of a particular new technology is to plot a technology S-curve for the technology of interest. The technology S-curve (referred to simply as an S-curve in the remainder of this review) is abundant in the information systems academic literature. In such instances, the S-curve typically plots some measure of performance against a cumulative measure of cost / effort. Figure 2.4 shows two hypothetical technology S-curves for technology products (S1

and S2), and a theoretical normal distribution for market adoption, and associated diffusion S-curve (S3) for market penetration (diffusion). As alluded to earlier, S3 essentially integrates the function of the normal market adoption distribution.

The normal distribution shown in Figure 2.4 is that derived by academic Everett Rogers, and described in detail in his book "Diffusion of Innovations" (Rogers, 2003). Rogers uses the S-curve in his explanation of how technology diffuses throughout a culture. The proportions of the five adopter categories under the distribution are determined by laying off standard deviations from the average time of adoption (see Rogers (2003), Figure 7-3).



Curve S3 shows how diffusion begins slowly, and after a critical mass is reached, diffusion is self sustaining and growth approximates exponential, gradually slowing until saturation is reached. Curves S1 and S2 represent technology curves. S1 is for an established technology which has reached the end of its life cycle, while curve S2 represents a disruptive technology directly competing against S1. S1 and S2 follow the same general pattern of all S-curves; initially improvements in performance per effort unit are small (simply because the newness of the technology makes it difficult for the organization to immediately identify any key performance drivers). After a period of sustained innovative effort, if the organization has identified the performance drivers, performance

improvements per effort unit increase greatly, and may approximate exponential. At some point however, the limits of the technology are reached, diminishing marginal returns set in, and performance gains per effort unit taper off, until ultimately, it is uneconomic to continue the cycle of innovation.

2.1.5.1 Technological Advances along the S-curve

As alluded to in section 2.1.1, technology typically advances along an S-curve in an incremental fashion and within an existing technology paradigm (Shane, 2008, p.23). Due to their experience working in the industry, incumbent producers, and not new market entrants make the majority (but not all) of technical progress (Christensen, 1992a). Incumbent producers can leverage extant technical and organizational capabilities to make improvements to underlying technology (Chandy & Tellis, 2000). They also possess the advantage of an existing customer base from which to identify market needs (Bower & Christensen, 1995). And finally, cash flow from the existing operation provides internal "free" private financing for innovation, putting them at a financial cost advantage compared to new entrants, when funding the investment in technology improvements (Teece, 1996).

2.1.5.2 Displaced Technology S-curves

When the limits of an existing technology are reached, the law of diminishing marginal returns ensues, and it becomes un-economic to pursue further technical improvements. At this point, if the market does not stagnate, new technologies are developed which challenge the current technology (Foster, 1985).

If the new technologies are not simple component changes, but represent architectural technology changes, they will initially underperform the existing technology in the market place when measured in terms of metrics most valued in the established market, so will most often initially be introduced into new market applications; thus making it difficult for existing technology producers to recognize the competitive threat of the new technology (Christensen, 1992). This concept of architectural technology changes inducing market innovations, is important and is discussed further by Tripsas (2007). Architectural technology changes also often re-define the market, and the metrics used to measure success. Ultimately, they have greater potential for performance enhancements than the existing technology; due to this, the newer technology is likely to surpass the existing technology in terms of metrics most valued in the established market. When this happens, customers will migrate to the new technology in great numbers, with the existing technology going into decline.

A classic example related to IT, is the introduction of the personal computer (microcomputers as they were referred to at the time) in the late 1970's, and the impact they had on Digital Equipment Corporation (often referred to as either DEC or Digital). During the 1960's through to the mid 1980's, Digital was one of the leading American vendors of computer systems (minicomputers), software, and peripherals, selling their products to the commercial sector. When PC's were introduced to the market, they targeted individual consumers (i.e. the "low end" of the market) as opposed to corporate, and were essentially ignored by Digital. The then CEO of Digital (Ken Olsen) is quoted at the time, as saying that there was no reason for individuals to have a computer in their home (Gatlin, 1999, p.15). Consequently, Digital failed to recognize the threat the PC represented to their business, and they failed to adopt the de facto standards that the fledgling PC industry was establishing. Eventually, Digital's corporate customers no longer wanted to pay the higher prices for Digital's highly proprietary (but better performing) computers, and started purchasing the much cheaper PCs, which adequately met their needs. By the early 1990's, Digital was losing its dominant market position, and in 1998, they were acquired by Compaq.

The introduction of new technologies into a market can be shown by the introduction of a new S-curve (curve S2 in figure 2.4).

It is important to identify those organizations that actually displace the S-curve. Typically, they tend to be new entrants (Christensen & Rosenbloom, 1995). Incumbents can (and do) innovate and develop new technologies capable of displacing the S-curve; but typically, they will persist with investing in existing technology (Brown, 1992). Only when their core products are threatened (diminishing their ability to serve key customers), or if they possess industry-specialized supporting assets (e.g. complementary assets, as discussed later in section 2.1.6), will the incumbent shift to the newer technology (Mitchell, 1989).

2.1.6 The Teece Model (Dynamic Capabilities Framework)

The models discussed up to this point were developed, in part, to provide insight into when it is most appropriate for organizations to innovate, and what circumstances are most supportive of innovation. The Dynamic Capabilities Framework extends this work, and improves the understanding of who will profit from innovation, and which organizations have the highest incentives to invest in particular innovations.

The Dynamic Capabilities Framework (Teece, 1996) was one of Teece's main contributions to the theory relating to technological evolution and innovation. This model demonstrated that two key factors influence what firms will profit from (and are more likely to invest in) particular innovations

arising in an environment of rapid technological change. These factors are the imitability of the innovation and complementary assets to the innovation. Imitability relates to the ease by which competitors can duplicate the technical innovation (or processes which underpin it), and complementary assets are assets used to produce and deliver new products and services.

Teece showed that if imitability is high and complementary assets are freely available, it will be difficult for the innovator to profit (i.e. receive economic rent) other than in the short run. If on the other hand, imitability is high and complementary assets are not freely available, then the organization holding the complementary assets will profit, regardless of who made the innovation. When imitability is low, and if the innovator also controls complementary assets, there is the potential to make high economic rent, while if imitability is low, but the innovator does not control the complementary assets, then the innovator will need to negotiate with the holder of complementary assets and share profit (Afuah, 2009). Figure 2.5 illustrates the importance of complementary assets to this model.

The Dynamic Capabilities Model also describes how organizations that continue to prosper in environments with high technological change, are typically those that are responsive to change, and can demonstrate timely and flexible product innovation, and an organizational ability co-ordinate and redeploy both internal and external (e.g. partners) competencies in the face of innovations in the market, especially those that have the potential to be disruptive (D. Teece & Pisano (1994); D. J. Teece, Pisano, & Shuen (1997)). The companies that prosper have greater dynamic capability; that is, they have a greater ability to understand what is needed from their organization at any given point in time, and apply their capabilities to the challenge at hand.

In terms of disruptive technology, when incumbents show a propensity to be responsive to change, it is possible that any potentially disruptive innovations (generated by incumbents or new market entrants), may be short-circuited by incumbents before their market position is adversely impacted, by adopting the technology, etc. It has also been demonstrated that when cooperative links between potential disruptors and incumbents exist to provide access to relevant complementary assets, there is a greater chance that a potentially disruptive technology will succeed without necessarily disrupting the market position of incumbents (Rothaermel, 2001).

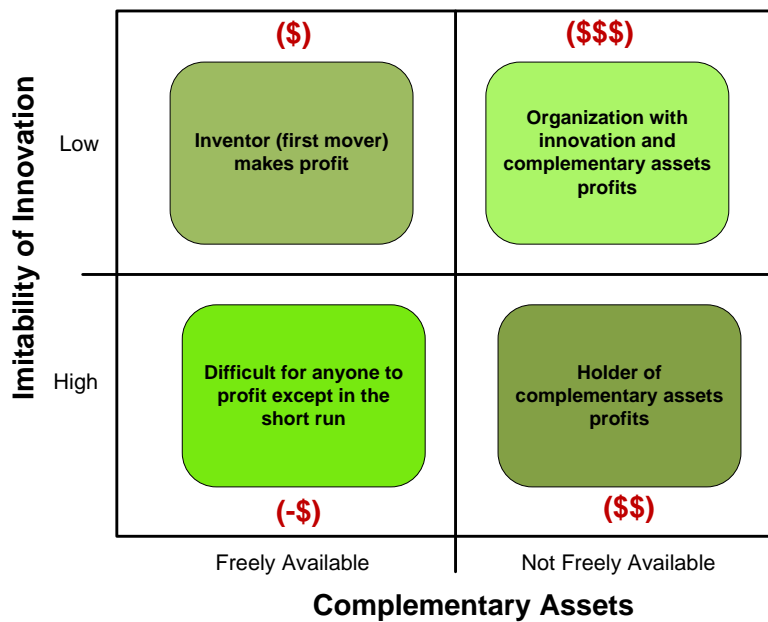


Figure 2.5 The Dynamic Capabilities Framework – importance of complementary assets.
(modified from Afuah (2009), p.125)

The Dynamic Capabilities Framework is the last static model discussed in this literature review. It should be pointed out that fundamentally, it builds on concepts that evolved in the static models discussed prior, but also from the Abernathy-Utterback model, a dynamic model that provides some useful insights into technology evolution. The Abernathy-Utterback model is discussed in the next section.

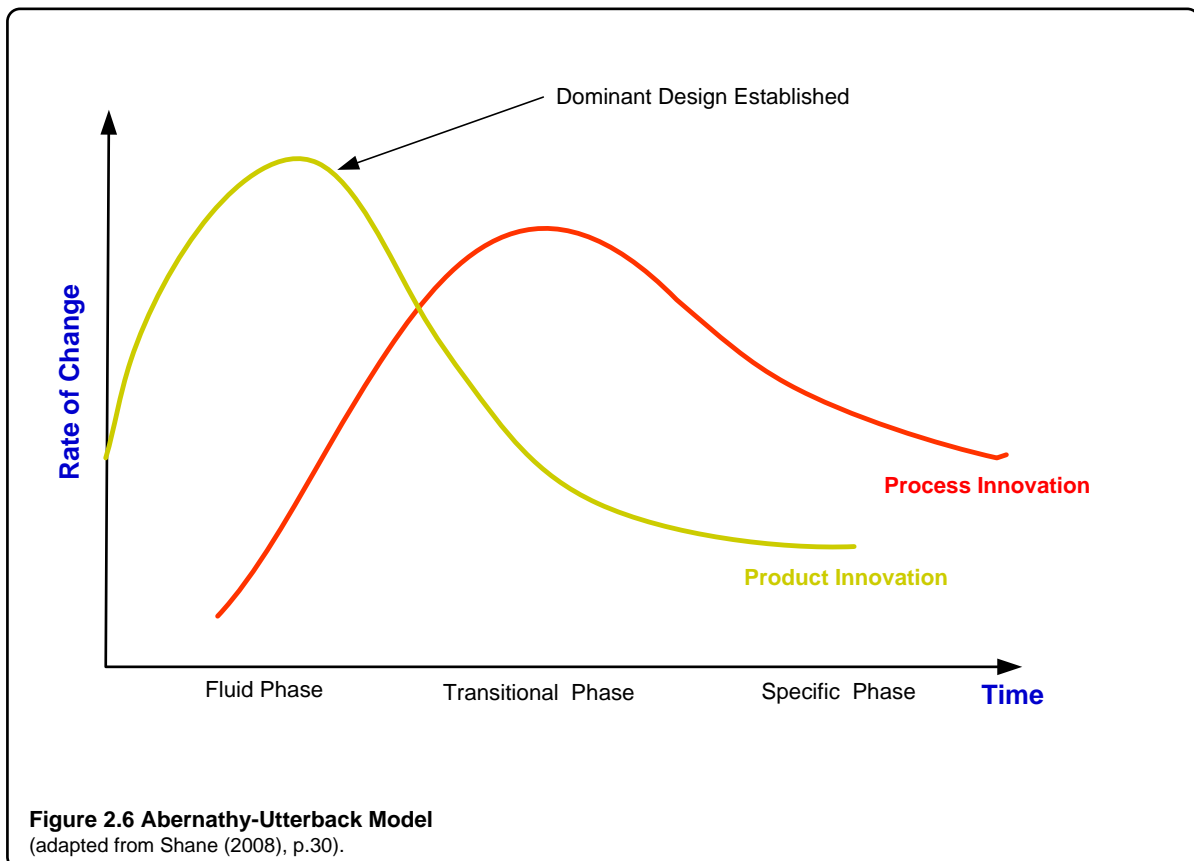
2.1.7 Abernathy-Utterback Model

Bill Abernathy and Jim Utterback developed a dynamic model of technology evolution. In this model, technology evolves through periods of incremental innovation, punctuated by periods of radical innovation. In this model, product innovation, process innovation, the competitive environment, and organizational structures all interact (Abernathy & Utterback, 1978).

Key to this model, are the concepts of product and process innovation. Product innovation is that innovation leading to the creation of new goods and services. Process innovation, is that which results in improvements in methods of creating or delivering products or services to market. (J. M. Utterback & Abernathy, 1975). The other key concept to the Abernathy-Utterback Model is that of a dominant design (Suarez & Utterback, 1995). The dominant design is the design whose main

characteristics and key components do not vary greatly between product models; it is also "the one that wins the allegiance of the marketplace, the one that competitors and innovators must adhere to if they hope to command significant market following" (J. Utterback, 1994, p.24).

Similar to the ideas underpinning the technology S-curve, Abernathy and Utterback proposed a three phase cycle for technology evolution; a fluid phase, a transitional phase, and a specific phase. This is shown in figure 2.6 below.



The fluid phase is the first phase of the model, and is characterized by a high degree of uncertainty. During this phase, the market is small, customer preferences are not clearly defined, and the diffusion of core technology is limited. Product innovation spurs the product differentiation by which market participants compete, while process innovation is low; production processes tend to be generic until the market converges on a dominant design (Afuah & Utterback, 1997). Once a dominant design is attained, the cycle is said to be in the transitional phase.

During the transitional phase, standardization around the dominant design begins to emerge, acceptance and diffusion of the technology starts to increase, and the market begins to grow.

Competitive advantage will be gained by the organization that most greatly influences the dominant design, as it gives them an opportunity to collect monopoly rents for a period.

During the specific phase of the model, the focus of market participants shifts from product differentiation, to one of process innovation, with the aim of reducing production costs (Gort & Klepper, 1982). This phase is often characterized by investment in specialized manufacturing equipment (Teece, 1986), and the number of market participants reduces, as larger organizations that can leverage economies of scale come to dominate the market (Klepper & Graddy, 1990). During this phase, competition is typical of an oligopoly; few competitors, intensely competing. The incumbents try to secure their position and gain competitive advantage through building relationships with suppliers, securing distribution channels, and trying to control complementary assets; all of which can create barriers to entry for new market entrants (Afuah & Utterback, 1997). Figure 2.7 below, maps the phases of the Abernathy-Utterback model to a standard technology S-curve. The salient points to note are that potentially disruptive innovations occur in the Fluid Phase, while the key inflection point on the curve, demarcating the change from the Fluid Phase to the Transition phase, is brought about by the emergence of the dominant design.

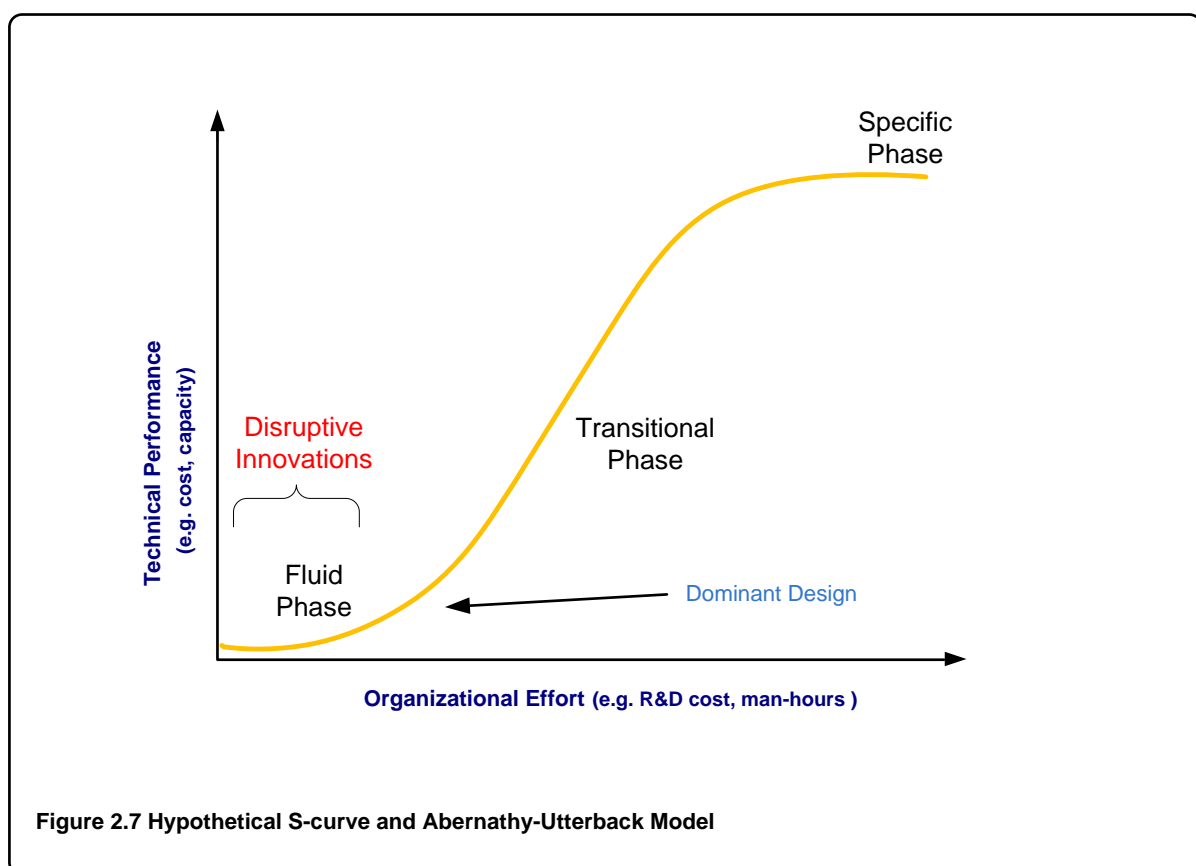


Figure 2.7 Hypothetical S-curve and Abernathy-Utterback Model

A number of important insights can be gleaned from the Abernathy-Utterback model. Firstly, the model shows (for technology intensive industries) that the fluid phase tends to support new firms, while the specific phase favors incumbent firms. The reasons for this arise from the fact that the dominant design has not yet been established. This situation provides new entrants the flexibility to not to have to use the same product design as incumbents, which levels the field; the prior knowledge of incumbents does not favor them (Suarez & Utterback, 1995). Additionally, not having settled on a dominant design, market participants operate on a small scale to minimize technical uncertainty (Mueller & Tilton, 1969), and also tend to have flat organizational structures (as discussed in section 2.1.5.2) which allows them to be more responsive.

The second insight from the Abernathy-Utterback model is that the number of firms in the market tends to change over the phases of the model. Typically, during the fluid phase, there is a large increase in the number of organizations in the market, due to high levels of new market entrants, with low exit rates. Soon after the dominant design has been established, market consolidation occurs, and the number of firms in the market drops to include only those organizations that are able to produce product consistent with the dominant design (Gort & Klepper, 1982). This 'shakeout' of the market is dramatic, with some authors demonstrating a drop in market participants averaging 52%; the range of the net decrease in firm numbers from peak values was from 4% to 87%, for a wide variety of products (Klepper & Graddy, 1990).

2.1.8 Abernathy-Utterback Model Extensions

The Abernathy-Utterback model does have some limitations. For example, it cannot predict patterns of industry evolution across all manufacturing industries with equal success. Assembled products are the industry class in which it works best; non-assembled products such as float-glass, is where the model performs weakest. It is thought that this results from the higher levels of process innovation associated with non-assembled products, and the fact that it is more expensive to protect process innovations from appropriation (Teece, 1986). It is also not as effective for the service industry (Barras, 1986).

A number of authors have described theories which extend the Abernathy-Utterback model to increase its predictive capability. Three such extensions are described below.

2.1.8.1 Reverse Product Cycle Model

Barras (1986) investigated whether the standard Abernathy-Utterback model could be applied to the service industry as successfully as it was applied to the manufacturing industry. He found that it could not, and proposed an alternate three stage model called the Reverse Product Cycle Theory.

According to this theory, in the first phase of the model, service industries tend to adopt new technologies that have been developed in the manufacturing industry first with the aim of increasing the efficiency (and thus reducing the costs) of existing services (Barras, 1986). Such process innovations are typically incremental (Damanpour & Gopalakrishnan, 2001). In the second stage of the model, technology is used to increase the effectiveness of services. Often, such innovations focus on improving the quality of customer service and its associated processes. In the final stage of the model, technology is used radically, to introduce new services (Damanpour & Gopalakrishnan, 2001).

Key differences between the Abernathy-Utterback model and the Reverse Product Cycle model include the timing of radical and incremental change; in the service industry the cycle of technological evolution begins with incremental innovation and goes on to radical innovation, whereas for the manufacturing industry, the cycle of technology evolution starts with radical innovation and ends with incremental innovation (Barras, 1986). Therefore, new firms have the initial advantage in the manufacturing industry before the appearance of the dominant design, while in the service industry, established firms have the initial advantage, as it reduces their costs of service delivery; in the service industry, the advantage shifts to new market entrants only after the technology has introduced new services, and reduced service costs permit them to enter the market (Barras, 1990).

2.1.8.2 Competence Enhancing and Destroying Innovations

Michael Tushman found that the Abernathy-Utterback model was unable to explain why some incumbent firms were able to transition to new technologies (i.e. shift to a new technology S-curve). Tushman & Anderson (1986) extended the Abernathy-Utterback model by introducing the concepts of competence-enhancing and competence-destroying innovations. Competence enhancing innovations can leverage existing organizational knowledge and skills, and other existing factors of production. Competence destroying innovations however, are unable to do the same, and may in fact frustrate an organization's existing factors of production.

When a new technology is competence-enhancing, incumbent firms will be prepared to invest in the technology, as they have the capability to successfully develop the technology. However, if the technology is competence destroying, the firm will tend not to attempt to transition to it, even if

they developed the technology themselves; the reasons for this include the fact that adopting competence destroying technology can be expensive, (Lynn, Morone, & Paulson, 1996), and also because their existing skill sets and organizational structures can make adopting a new technology difficult (Afuah & Utterback, 1997). Such competence destroying technologies bear a strong resemblance to disruptive innovation.

2.1.8.3 New Market Entrant Displacement of S-Curve.

There are a number of reasons why new entrants are typically the market participant that displaces the S-curve (Shane, 2008: pp.25-26):

- Incumbents have no incentive to introduce new technology; typically, the performance of new technology under-performs the existing technology.
- Incumbents have significantly invested in the existing technology, and want to maximize return and minimize write-offs from such investment.
- Products based on new technology have the potential to cannibalize sales of existing products based on the current technology. Any extra revenue derived from the new technology erodes revenue from the existing product base; that is, there is the possibility that costs will rise while net revenues decrease.
- Managers within incumbent organizations often fail to recognize the new technology as a threat, due to embedded organizational routines leading managers to "filter out" information relating to new technologies, and to view them negatively.
- Incumbent firms will compete and initially improve the performance of existing technologies in response to new technology.
- Organizations often mold their structure and align internal processes to fit comfortably with their technology choice. This structural alignment can make it difficult to adopt new technologies, as it will require concomitant changes in organizational structures.

2.1.9 Disruptive Innovation Theory

Disruptive innovation, is an important mechanism which can assist organizations to broaden existing and develop new markets, and can provide new functionality, which may disrupt existing market linkages (D. Yu & Hang, 2010). Clayton Christensen was one of the first authors to make reference to disruptive technology in his seminal book "The Innovator's Dilemma" (Christensen, 1997). In this book, Christensen describes the dichotomy that exists between 'sustaining' and 'disruptive'

technologies. By his second book "The Innovator's Solution" (Christensen & Raynor, 2003), Christensen had replaced the term 'disruptive technology' with 'disruptive innovation', arguing that technology *per se* was not itself disruptive, but rather, it was the use of the technology, and the innovation it supported, which was disruptive to the market structure.

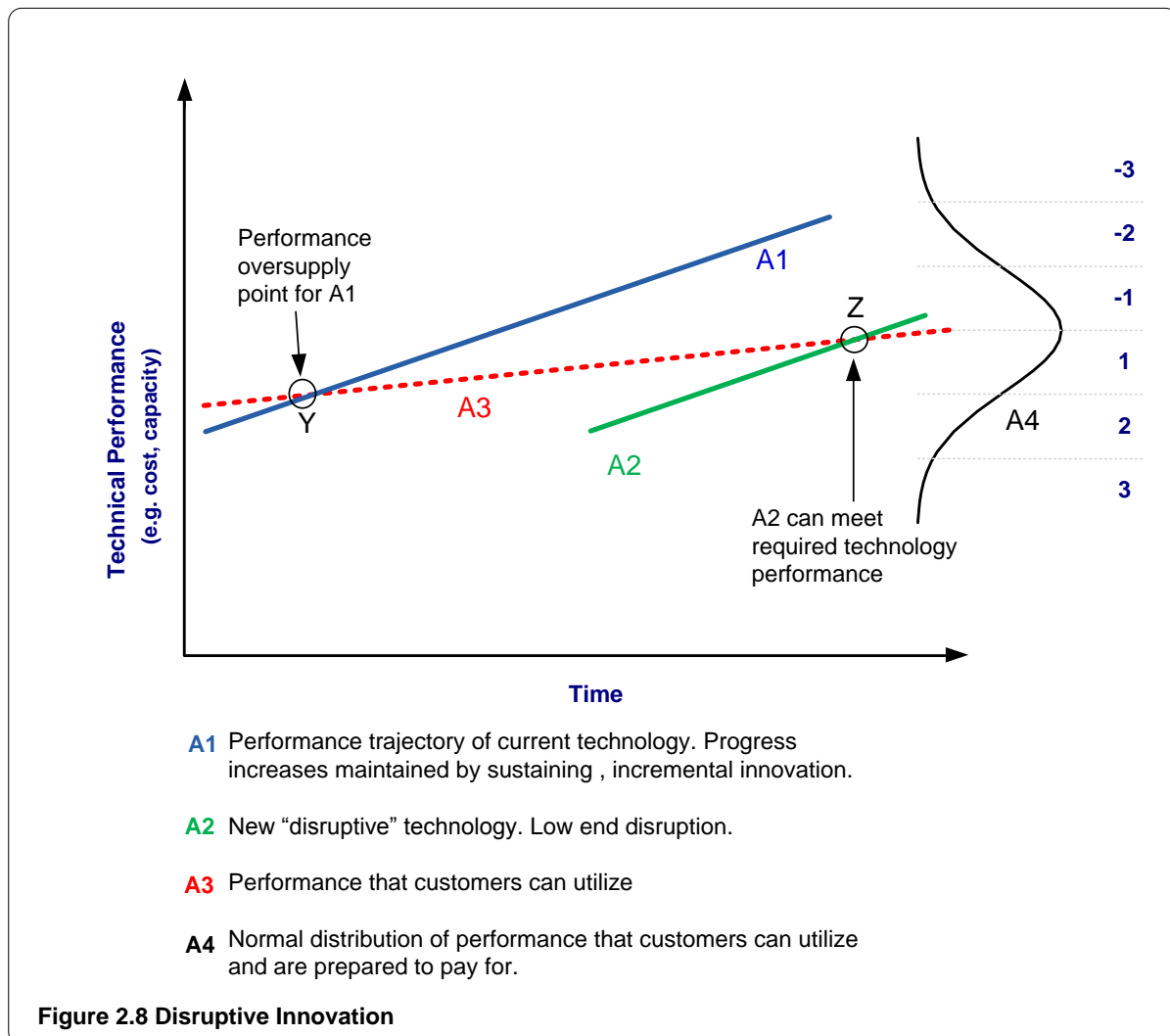
Important to understanding Christensen's model of disruptive innovation, is the concept of technology substitution. By understanding when/how different technologies are adopted, you can gain insight as to when new technologies may be introduced into a particular market context, and identify those which may be disruptive (Paap & Katz, 2004). Paap and Katz also point out that before technology substitution can occur, there needs to be both an unmet need relating to dominant factors influencing business performance, and an inability of the current technology to address the need in a competitive manner. The three drivers of technological substitution as described by Paap and Katz are:

- Where the existing technology matures relative to the dominant business driver (i.e. it becomes less relevant).
- Where the dominant business driver evolves to take on a different form to which the existing technology is not well matched.
- Where the business environment changes and a new business driver emerges, again to which the existing technology is not well matched.

In "The Innovator's Dilemma", Christensen described how sustaining innovations (i.e. incremental innovations) were those developments that allowed organizations to make marginal changes requiring gradual change to processes, and which increased the product performance along dimensions that were important to their key customer base. Disruptive innovations by contrast, were those innovations which were radical in nature, often driven by unexpected technological breakthroughs. Such innovations were often not initially valued by key market segments, and typically appeared to be cheaper, simpler, and often possessed inferior characteristics when compared to the existing products (although valued by some niche market segment). However, the disruptive innovation ultimately overturned existing products and markets.

The dilemma Christensen refers to, is that which organizations face when deciding whether to embrace or ignore innovative developments that do not appear to address their customer's needs. Such conflicting challenges of dualism, where successful firms are faced with the task of sustaining the success of their business today, while at the same time, embracing potentially disruptive innovations to try and sustain their competitive edge for tomorrow, has been referred to as the "tyranny of success" (Paap and Katz 2004).

In his model of disruptive innovation, Christensen (1997) describes how the decision to ignore such innovative technologies when they are substitutable and their technological trajectories interact with that of the incumbent technology, can prove fatal to the incumbent firm. Figure 2.8 illustrates Christensen’s model of disruptive innovation.



In figure 2.8, it can be seen that as the performance required (and can be utilized) by the customers of the incumbent technology increases over time (A3), so does the performance trajectory of the incumbent technology (A1). Often, the trajectory of the incumbent technology is different from that of the customer’s performance requirement. If the trajectory slopes are different, and the supplied performance of the incumbent technology exceeds that which is demanded and can be utilized by the customers (from point Y in the figure), then a substitutable technology that may have only been competitive in related niche markets can migrate into the incumbent’s customer network. But only if its performance trajectory (A2) meets the customer performance needs (as it does at point Z in the figure). It is by this path that innovations can enter new markets.

One fact that interested Christensen was that incumbent firms were often seemingly unable to adapt to the radical technology presented by the disruptive innovation (despite sometimes being responsible for the initial technological breakthrough themselves). Christensen proposed that this was due to the fact that the key customers of the incumbent firms were unwilling to move to the new technology, because it was perceived as inferior to the existing technology on some metric important to them (Christensen, 1997, p.15). It is this perceived lack of performance relative to the incumbent technology, which leads to such disruptions being labeled *ex post* as “low-end” or “attacks from below”, as in A2 in figure 2.8¹. Key customers would thus pressure the incumbent firm to focus on incremental improvements based on existing technology. It is this concept of pressure from key customers which Christensen encapsulates in his concept of a firm's value network (C. M. Christensen, 1997, p.35), which is essentially the context within which a firm identifies and responds to key customer needs, reacts to competitors, and strives for profit. Due to the value network, incumbent firms may pay too much attention to powerful key customers, fail to identify disruptive technology early, or simply decide to wait until the market niche served by the disruptive innovation is large enough to be profitable. Such value networks can embed themselves in a firm in numerous ways, such as organizational structure, competitive and R&D strategies, interactions with customers, market strategies, etc; all with the intent of optimizing the performance trajectory of their product. As referred to earlier, this can result in considerable inertia accumulating within the trajectory the organization is following, making it difficult for them to follow a different path.

To counter potentially negative impacts of the value network in dealing with disruptive technology, Christensen provides examples of where incumbents have spawned independent companies to commercialize disruptive technologies (Christensen, 1997, p.104).

Christensen also points out that while potentially disruptive innovations often start off competing with the incumbent at a disadvantage in terms of metrics key customers are demanding, they also possess characteristics along other dimensions which far outperform that offered by the incumbent. If the potentially disruptive innovation starts to meet the minimum technology performance required by key customers, then the innovation can start competing directly with the incumbent along the dimensions of the current metrics, but also along other metrics, and can thus begin to subtly move key business drivers along metrics where it has clear competitive advantage. If it is able to do this, it can diminish the relevance of the incumbent, and disrupt the market structure; i.e. it becomes a disruptive technology.

¹ Christensen & Raynor (2003) also describe new-market disruptive innovations, which are those which create an entirely new value network which do not compete directly with incumbent offerings. In such cases, the innovator has to deal with non-consumption - i.e. they have to establish a new market.

In research that was complementary to Christensen's, Veryzer (1998) demonstrated that the perceived value of what were essentially substitutable products or services, determined whether or not they were discontinuous. Veryzer identified that firms can deliver 3 types of discontinuity, as perceived by their customers:

1. When technological capability is advanced, while product capability remains the same, a technological discontinuity is perceived.
2. When product capability is advanced, while technological capability remains the same, a commercial discontinuity is perceived.
3. When both product and technological capability are advanced, then a technological and product discontinuity is perceived.

2.1.9.1 Disruptive Innovation: An Ongoing Debate

While Christensen's original ideas still have wide support, there are some researchers who argue that Christensen's definition of disruptive innovation is vague (Danneels, 2004). Some researchers propose that elements of Christensen's theory are wrong; (Tellis, 2006) suggests that it is difficult to discriminate between a low-end potentially disruptive innovation, and an inferior technology which would not go on to be disruptive.

Others adopt Christensen's main ideas with their own refinements. For example, (Adner, 2002) proposes that from a consumer perspective, marginal utility of successive performance gains in key performance dimensions in addition to Christensen's value propositions and lower price, explain consumers' choice to switch from the incumbent technology – a sustaining technology paradigm, to a disruptive innovation. He goes on to say that after their functional requirements are satisfied, a customer's focus shifts to the differences in absolute price rather than differences in price/performance points, which substantiates Christensen's concept of low end disruptions. Govindarajan & Kopalle (2006) extend Christensen's model of low-end disruptive innovations (inferior performance / low price), to include what they term high-end disruptive innovations (inferior performance / high price). They cite mobile phones as an example, which when originally introduced to the market, had inferior performance when compared to land-line phones and were expensive, but some consumers valued their convenience and portability. Over time, the performance of mobile phones increased to the point where they could compete with land-line phones on main consumer performance dimensions (and exceed them on many others), and their price-point dropped to one which mainstream consumers were prepared to pay, and they disrupted the market. Govindarajan & Kopalle's concept was further extended by Schmidt & Druehl (2008), who introduced the concept of a 'new market' disruption, where such disruptions could be fringe-

market, low end encroachment, or detached market, low end encroachment. In this framework, cell phones would be considered to be a detached market, low end encroachment.

Govindarajan & Kopalle (2006) also determined (and verified by rigorous statistical analyses) those characteristics most strongly associated with a disruptive innovation. These characteristics were that the innovation:

- Possessed inferior performance characteristics on measures important to mainstream customers
- Offered a different proposition to attract either a new customer segment or a more price sensitive mainstream customer segment
- Was offered at a lower price-point
- Enter the market from a niche position and migrate into the mainstream market

Yu & Hang (2010) also highlight what they believe are three key features often overlooked or misunderstood relating to disruptive innovations. First, disruption is relative; what may be disruptive to one organization may not be disruptive to another. They cite an example provided by Christensen & Raynor (2003, pp.41-42); Dell originally sold its PCs via mail-order or telephone, and moved to the now familiar internet selling model. For Dell, this was a sustaining innovation. However, for Compaq and IBM for example, who relied on their retail partners for sales, marketing directly to consumers via the internet was disruptive.

Second, disruptive innovation does not have to mean that incumbents have to be replaced by new entrants, and in some cases can actually be the disruptors themselves. Yu & Hang (2010) give the example of how incumbents can focus at the high end of their market (where price sensitivity is least) to negate the impact of a potentially disruptive innovation, and also of Intel's success with the Celeron chip-set, as a description of how an incumbent introduced a potentially disruptive innovation, which had a significant impact on an existing market, without displacing it. In this case, Intel's Chairman, predicted that the speed of Intel's processors was going to overshoot the speed that mainstream users could utilize, and that the processor speeds of competitors at the low end of the market (e.g. AMD) would increase to the point where they could invade the core of Intel's market. In response, Intel set up a separate company in Israel, to produce the Celeron chip, which subsequently went on to become their biggest selling chip, and potentially prevented AMD from disrupting their market (C. M. Christensen, 2006).

Third, destructive innovations do not have to be disruptive. For example, innovations, which have high performance characteristics, but a low cost and price-point, can have destructive impacts on incumbents, without disrupting the market. Yu & Hang (2010) cite the example of IBM's SiGe-based

chips. These chips significantly out-performed existing communication chips by a factor of 4, but were able to be manufactured using existing modified chip manufacturing plants (Haramé & Meyerson, 2001). Due to this fact, other key incumbents in the market (Alcatel-Lucent, Motorola), were able to follow IBM's lead and adopt the potentially disruptive innovation (at considerable but not insurmountable cost to themselves) and maintain their market incumbency.

2.1.9.1.1 Disruptive Innovation Defined

As alluded to in prior sections, there is a wide view of what constitutes a disruptive innovation, with many authors adopting a "point solution" mentality, with definitions being created to address specific concerns they may have, as opposed to providing a more holistic definition. Numerous authors identify a lack of clear definition as being an ongoing issue in research relating to disruptive innovation (e.g. Yu & Hang (2010) ; Kostoff, Boylan, & Simons (2004)).

For the purposes of this project, the working definition of disruptive technologies based on that of Thomond, Herzberg, & Lettice (2003, p.6) is that they are:

Processes, technologies (products or services), or business models that enable an organization to action significant change in competitive pressures, and transform the demand and needs of an extant market. As a result, new competing performance criteria may be introduced, incompatible with the incumbent technology, which allows the disruptor to capture market share; often in an under-served customer segment, as the key customer segment initially undervalues the potentially disruptive innovation. As the disruption becomes established, path and resource dependence of the incumbent technology/organization, results in the incumbent losing their prominent market position, and the disruptor gaining significant market share.

Note that this study considers radical and breakthrough technologies as synonyms for disruptive technologies.

The key components of this definition are that:

- It is not technology-bound (it also includes processes, products and services, and business models).
- It highlights that dominant business drivers may be changed by a disruption.
- It highlights that the disruptive innovation often starts out serving an under-served / niche segment, as the key market segment undervalues the innovation.
- The incumbent often loses their preeminent market position after some time.
- The disruptor typically dominates the market after a while.

Using this definition, if an incumbent switched to a potentially disruptive innovation and retained market share, the new innovation would not be viewed as disruptive, as no market disruption occurred. In the vernacular of Veryzer (1998), a technical discontinuity will have been seen to occur, but not a commercial discontinuity (i.e. market disruption).

The above definition will be used as the basis for answering the research question of this project; i.e. can discriminatory linguistic markers be found for disruptive innovation. It also feeds into the following section.

2.1.9.1.2 *Ex Ante Identification of Disruptive Innovations*

Researchers have questioned the predictive ability of the theory around disruptive innovation. Arguments range from the idea that some organizations are simply lucky in their choice of technology (Barney, 1997, p.17), through to the fact that making *ex ante* predictions predicated on determining what performance a market will demand in the future is fraught with difficulty, as no widely accepted methods exist for making such determinations (Danneels, 2004). Tellis (2006) makes the point that if it is only possible to identify disruptive innovations *post hoc*, there is little if any predictive value in the theory.

However, numerous authors argue that it should be possible to identify disruptive innovations *ex ante*. Danneels (2004) for example, suggests that extant methods used for technology forecasting may form the basis of some future technique; Schmidt (2004) suggests using 'part-worth' curves to look for instances where incumbent technologies are over-serving key customers' primary needs while under-serving their secondary needs - such instances will make the market amenable to a lower-cost disruptive product entering, if it can effectively compete on serving the secondary needs of key customers; Govindarajan & Kopalle (2006) suggest that using their disruptive innovation framework it is possible to identify which organizations may succeed in commercializing disruptive innovations; Druehl & Schmidt (2008) suggest that their encroachment framework, especially the indicators of strategies based upon a "willingness to cannibalize", can be used as an indicator of which organizations are more likely to develop successful disruptive innovations; and Yu & Hang (2010) suggest that an organization's prior success or failure in developing or commercializing disruptive innovations are key indicators of their future success in such endeavors.

A key assumption of this project is that *ex ante* identification of disruptive innovations is possible. Tables 2.1 and 2.2 below give some of the characteristics of disruptive innovations and supporting environmental conditions of such innovations, through the lens of the various models discussed up

to this point. Such characteristics will aid the search for predictive linguistic markers for disruptive innovation as described in later sections.

As will be discussed in section 3.2.1 (pp. 43-45), attribute set categories consisting of key words from the practitioner discourse literature were used to capture reference to key themes within the literature. The categories used were Benefit (B), Consumer (C), External Environment (E), Feature (F), Industry Analyst (IA), Industry (I), Output (O), Task (T), Technology (TY), and Vendor (V). The 'Cat.' Column in tables 2.1 and 2.2 references these categories, linking theory to the attribute sets derived. The ID column supports referencing back to these tables in later sections of the document.

Table 2.1. Characteristics of Disruptive Innovations.

Characteristic of Disruptive Innovation	Model / Author	Cat.	ID
Involve revolutionary/radical changes in technology.	Dewar & Dutton (1986)	B, F, IA	1
Involve significant departure from existing practice.	Dewar & Dutton (1986)	B, IA	2
Are often high-risk, high-return developments.	Shane (2008)	B	3
Establish a new dominant design, and a new set of core design concepts.	Henderson & Clark (1990)	B	4
Present challenges to incumbent, both in terms of instigating, and recognizing.	Henderson & Clark (1990)	O, TY	5
Usually underperform the incumbent technology, using measures most valued in existing market.	Christensen (1992)		6
Often introduced to market in new market applications.	Christensen (1992)	B, F	7
Redefine the market, and the metrics used to measure success.	Christensen (1992)	B, E	8
Possess greater potential for performance enhancements than incumbent technology.	Christensen (1992)		9
Often introduced by new market entrants.	Christensen & Rosenbloom (1995)	C,I,V	10
Arise as a result of product innovation.	Abernathy-Utterback (1997)	B	11
Often viewed by key market segments as cheaper, simpler, and inferior.	Christensen (1992)		12
At some point, they will meet a minimum level of performance required by the key market segment.	Christensen (1992)	B	13
If developed by an incumbent, they will often spawn a new company to commercialize the innovation, or some other autonomous business structure - e.g. a "skunk works".	C.M. Christensen (1997); C. M. Christensen (2006); Chesbrough & Crowther (2006)	V	14
The disruptor not the incumbent will eventually dominate the market.	Christensen (1992)	E	15

A number of environmental factors are often seen before a potentially disruptive innovation can disrupt the market; i.e. become a disruptive innovation (consistent with the definition used for this study). Table 2.2 following lists some of the key factors:

Table 2.2. Environmental Conditions Indicative of Disruptive Innovations.

Environmental Conditions Indicative of Disruptive Innovation	Model / Author	Cat.	ID
The limits of the existing technology are reached, and it becomes uneconomic to pursue further technical improvements.	Foster (1985)	T	16
Incumbents are unresponsive to change (due to structural alignment within the organization to the technology being produced, or due to a dominant design being established).	Teece (1996)	O	17
The market is undergoing a period of product innovation.	Abernathy-Utterback	V	18
There is a high degree of uncertainty in the market, with customer preferences not clearly defined.	Abernathy-Utterback	F,I,TY	19
There are many small 'vendors' (indicating that the market is in the fluid phase, when disruptive innovations typically arise), or an oligopoly exists with the incumbent product, which indicates that the market is in the specific phase (and so may be about to enter a new cycle - i.e. enter the fluid phase).	Abernathy-Utterback	I,V	20
Technological substitutes do exist in the market <i>and</i> an unmet need exists relating to key business drivers which the incumbent technology cannot meet in a competitive manner.	Paap & Katz (2004)		21
The incumbent technology provides more performance than key market segments can utilize.	Christensen (1992)	B	22
The disruptor organization has a separate corporate team responsible for assessing and implementing disruptive innovation ideas.	Christensen & Raynor (2003)	V	23
The disruptor organization is run by its founders (as opposed to professional managers).	Yu & Hang (2010)		24
Frustrated engineers from an incumbent organization establish a company to introduce potentially disruptive technologies.	C. M. Christensen & Bower (1996)	V	25
The disruptor organization has cultural elements which emphasize entrepreneurship, creativity, flexibility, and risk-taking.	Govindarajan & Kopalle (2006)		26
Structured routines relating to using financial returns or traditional market research to assess potential disruptive technologies are not deeply embedded within the disruptor organization.	Nelson & Winter (1977)		27
Any resource dependencies that could inhibit the exploitation of disruptive innovations by causing the organization to re-prioritize investment spending to those opportunities in which they have accumulated resources are not present in the disruptor organization.	C. M. Christensen, (2006); Christensen & Bower (1996)		28
Smaller organizations are associated with more successful R&D initiatives.	Lee & Chen (2009)	E,V	29
Organizational size correlates negatively with successful commercialization of disruptive innovations.	DeTienne & Koberg (2002)	V	30
Forms of cooperative links between potential disruptors and incumbents exist (to provide access to relevant complementary assets). Such links are associated with successful disruptive innovations.	Teece; Rothaermel (2001); Claude-gaudillat & Quélin (2006);		31
The disruptor organization is publically owned and located within a regional technology cluster (this increases their attractiveness as an alliance partner, providing benefits alluded to in the point above).	Rothaermel (2002)		32
Formal relationships exist between the potential disruptor organization and technology suppliers and partners.	Myers, Sumpter, & Walsh (2002)		33
The potential disruptor organization has a higher emerging customer orientation than the orientation for their existing customer base.	Govindarajan & Kopalle (2004)	B,E	34

The concepts in the tables above are relatively generic, and will be used to form that basis of the search for discriminatory linguistic markers for disruptive innovation, using SaaS BI as a case. The following section provides a high level background relating to Software as a Service (SaaS), business intelligence and SaaS BI.

2.2 Software as a Service

Software as a Service (SaaS) began to gain acceptance in the information technology market by 2000 (Bennett et al., 2000). At its most basic level, the model for software delivery via SaaS involves a software company presenting a single copy of their software to the Internet (using their own dedicated servers), with individual organizations then "renting" access to the software via a subscription based model. Models where multiple organizations can use the same infrastructure are referred to as multi-tenant architecture models (Bezemer & Zaidman, 2010). Note that while the infrastructure is physically shared between tenants, at the logical level at least, the infrastructure is perceived by the individual organizations as being uniquely theirs.

In this model, the software company providing the service operates, maintains and supports its customers via a centralized service. From a consumer perspective, the benefits of SaaS are a lower total-cost-of-ownership, and improved return-on-investment. The drivers of these benefits are the facts that:

- Tenants do not have the up-front cost of purchasing software and hardware.
- Economies of scale afforded to the SaaS provider by a multi-tenancy, virtualized, hosted infrastructure, allow the provider to offer maintenance etc, at a rate lower than would be achievable by the tenant should they have an on-premises solution.
- Subscription based services align costs with actual usage, cost typically rising with increasing usage.

It is acknowledged that there is some confusion between the concept of cloud computing and SaaS. For the purposes of this study, the term SaaS is used to be consistent with the definition of cloud computing provided by Armbrust et al., (2010), who refer to cloud computing as both the applications delivered as services over the internet, and the hardware and systems software in the datacenters that provide those services.

2.3 Business Intelligence

Business Intelligence as a concept, can be traced back to the 1958 paper authored by IBM computer scientist Hans Luhn, who defined business intelligence as "... a collection of activities carried on for

whatever purpose, be it science, technology, commerce, industry, law, government, defense, et cetera. The communication facility serving the conduct of a business (in the broad sense) may be referred to as an intelligence system. The notion of intelligence is also defined here, in a more general sense, as the ability to apprehend the interrelationships of presented facts in such a way as to guide action towards a desired goal" (in Chee et al., 2009).

A more recent definition is provided by Howson (2008, p.2), who defines business intelligence as a set of technologies and processes that "allow people at all levels of an organization to access, interact with, and analyze data to manage the business, improve performance, discover opportunities, and operate efficiently".

Davenport & Harris (2007, p.12) have a slightly wider definition which includes the processes around the collection and management of data, and also the analytical and computational activities undertaken on the data (for its conversion into actionable information). Davis, Miller, & Russell (2006) also emphasize the critical importance of analytics, to the concept of business intelligence. For the purposes of this study, a broad definition of business intelligence is used, consistent with the definitions above.

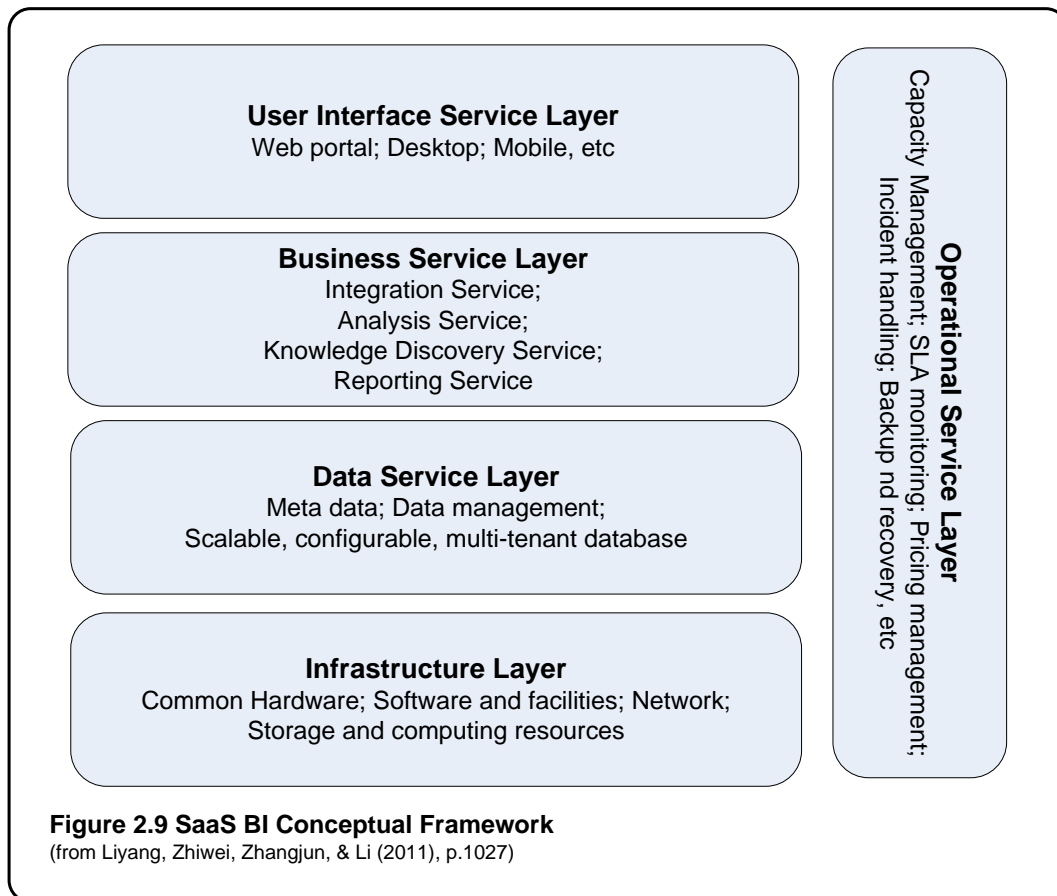
From this point on, the acronym BI is used to refer to business intelligence. It is also assumed that as the concept of BI is mature, any vendors describing their technology as BI are referring to BI as described above.

2.4 SaaS BI

SaaS BI is a new and nascent technology. In this study, the framework for SaaS BI provided by Liyang, Zhiwei, Zhangjun, & Li (2011) is used to define the scope of what is considered to be SaaS BI applications. The framework of Liyang, Zhiwei, Zhangjun, & Li (2011) is illustrated in figure 2.9.

In the figure, the *infrastructure layer* is used to provide common hardware, software and facilities (including network, storage and computing resources). This layer would typically be hosted by SaaS vendors, and be transparent to users. The infrastructure provided at this level is required to be highly available, and scalable. The *data service layer*, consists of database systems, as well as meta data and data management systems. A multi-tenant data architecture is critical to this layer. The *business service layer* consists of four key services which provide the four key processes usually associated with business intelligence (integration, analytics, knowledge discovery, and reporting). The *user interface services layer* provides the data access layer usually associated with on-premises BI systems, and provides different interfaces to facilitate user access to the BI application. Typically

the components in this layer provide the ability for developers (from the SaaS BI provider, or the consuming organization) to create web pages using utility Web technologies; allow companies to assemble user interface elements into a new user interface, and logic control, supporting the direction of user interface behavior. Finally, the *operational service layer* provides a management layer for the operation of the SaaS BI system, ensuring service delivered is in accordance to that described in the SLAs (service level agreements).



The 2011 Gartner Hype-Cycle for Business Intelligence (Gartner, 2011), describes SaaS BI as an emerging technology, with a market penetration of 1 to 5%. Figure 2.10 following, shows the Gartner Hype Cycle for Business Intelligence (for 2011).

In terms of the Gartner Hype-Cycle, SaaS BI is still a growing market, with SaaS BI passing the 'Peak of Inflated Expectations' (see figure 2.10), indicating that there are a large number of independent offerings, and also an acceptance by consumers for this form of pricing and delivery model.

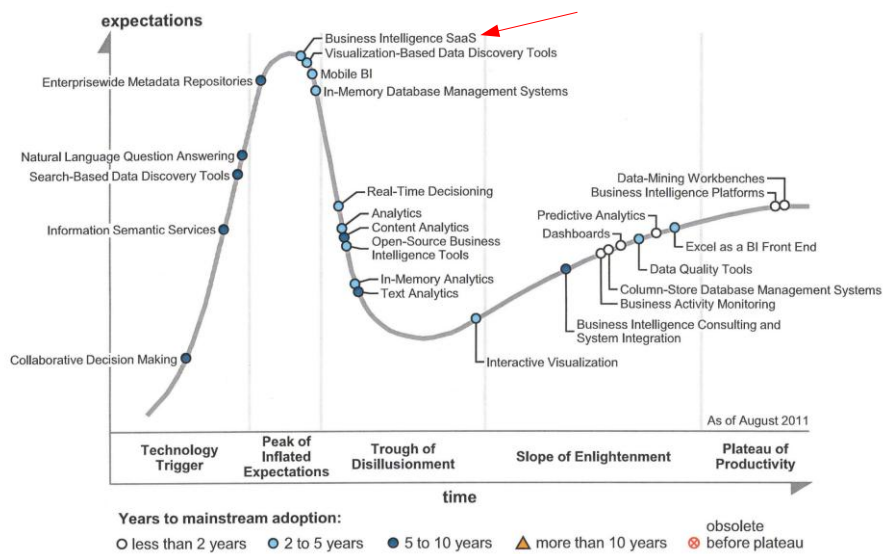


Figure 2.10 Gartner Hype Cycle for Business Intelligence, 2011
(from Gartner (2011), p.7)

Gartner (2011) advise organizations with simpler requirements to consider the use of SaaS BI, explaining that SaaS BI implementations tend to perform better for narrowly defined applications as opposed to broad deployments (e.g., replacing an entire BI/data warehouse infrastructure). Organizations with more complex requirements should evaluate traditional vendor offerings, and consider consuming on-premises software, but on a hosted basis. On-demand BI poses integration challenges for enterprises that need to export data to and extract data (and metadata) from the service provider for incorporation into their BI infrastructures however.

Gartner describes a key advantage of SaaS BI as giving consumers the ability to deploy one or more components of BI without significant IT involvement or the need to deploy and maintain an on-premises BI solution. The same report also lists the drivers of SaaS BI adoption as being reduced startup costs, easier support with minimal IT contribution, and easier deployments. Gartner did not think that sophisticated functionality was driving adoption, although does make mention of organizational frustration with long deployment cycles, high cost, complicated upgrade processes, and the IT infrastructure demanded by traditional BI solutions. SaaS BI as an alternative, Gartner goes on to say, offers quick, low-cost and more easily deployable solutions, that have proved to be popular, particularly in SMBs (small to medium businesses) and also in the individual departments of large organizations that lack or do not want to support internal BI resources.

Other authors have described additional benefits associated with SaaS BI, such as the freeing up of internal resource to focus on core competencies; “anytime / anywhere” access, provided users have an internet connection; more technology choices; low entry barriers mean new application types can easily be trialed; faster product cycles; and transparent maintenance updates (Ouf et al., 2011).

It should be noted that some large firms are also gearing up to provide BI and Analytics as a service; for example, Deloitte purchased Oco in May 2011. Oco is an enterprise-class software as a service (SaaS) business analytics solution. The fact that one of the "Big 5" consultancy firms has added a SaaS BI product to its service portfolio may be an indicator that the technology is disruptive, and we may be witnessing the beginning of a market disruption.

Numerous vendors offer products in the SaaS BI space. Some of the main vendors include (but are not limited to) SAP BusinessObjects, SAS, IBM Cognos Now!, Oracle, PivotLink, Oco, Dimensional Insight OnDemandIQ, Birst, GoodData, Pervasive Software, Antenna Software, Cloud9 Analytics, CVM Solutions, Exoprise Systems, Gagen, Host Analytics, KnowledgeTree, LiveOps, Reval, and Taleo.

3 Methodology

This study utilizes content analysis to search for discriminative linguistic markers within ICT practitioner discourse that would support the ex ante classification of SaaS BI as a disruptive technology. The following sections describe the nature and purpose of content analysis, and provide detail of a process for undertaking content analysis.

3.1 Nature and Purpose of Content Analysis

Content analysis consists of a wide variety of theoretical frameworks, methodologies, and analytical techniques. Stemler (2001, p.1) defined content analysis as “a systematic, replicable technique for compressing many words of text into fewer content categories based upon explicit rules of coding”. Neuendorf (2011, p.277) described content analysis as a “summarizing, quantitative analysis of messages that relies on the scientific method, including attention to objectivity/inter-subjectivity, replicability, and hypothesis testing. It is not limited as to the types of messages that may be analyzed, nor as to the types of variables that might be measured”. And finally, Duriau, Reger, & Pfarrer (2007, p.6) provide a definition of content analysis based upon earlier works, that is both minimal and all-encompassing: "Content analysis is any methodological measurement applied to text (or other symbolic materials) for social science purposes". The three above definitions provide the conceptual framework for what is considered to be content analysis for the purposes of this project.

Duriau, Reger, & Pfarrer (2007) also discuss the importance of language in human cognition, and how this fact underpins the value of content analysis as a research methodology. It is this role of language in human cognition that forms one of the basic assumptions of content analysis; namely that analysis of texts provides researchers with insight into the cognitive schemas of the authors of those texts - i.e. insight into what they were thinking at the time of writing (Gephart, 1993).

Word frequency is seen as the most basic form of content analysis, and can be considered to be indicative of cognitive centrality (Duriau, Reger, & Pfarrer 2007), and importance (Abrahamson & Hambrick, 1997). It is also assumed that change in word usage reflects a change in cognitive schema, and also, that particular groupings of words intimate latent themes, and that co-occurrence of words reflects associations between latent themes (Duriau, Reger, & Pfarrer 2007). When content analysis stops at identifying and quantifying particular words or content in text, for the purpose of understanding the contextual use of the content, the analysis is said to be quantitative or manifest content analysis. If however, latent analysis is undertaken, whereby an interpretive process is undertaken of the content, to discover underlying meaning, then the analysis is said to be qualitative, or summative content analysis (Hsieh & Shannon, 2005).

Many authors have demonstrated key benefits of content analysis in social science research, particularly in the organizational context. For example, the technique provides a replicable method to identify 'deep' individual or collective structures, such as values, intentions, attitudes, and cognitive schemas (Carley, 1997). Additionally, content analysis provides analytical flexibility, allowing analysis of content to occur at two levels; at the manifest level of word frequencies and 'symbol coincidences', and also at the level of latent content - deeper meaning embodied in the text and thematic interrelationships (Woodrum, 1984). Content analysis also supports inductive and deductive research (Roberts, 1989).

As alluded to earlier, another advantage often cited regarding content analysis in the organizational setting, is that it supports longitudinal research designs, due to the availability of comparable corporate information over time, via annual reports, trade magazines, etc (Kabanoff & Holt, 1996).

The advent of computer aided textual analysis greatly increased the scalability of content analysis by providing enhanced access to textual information in electronic format, and via task automation (e.g. data storage, dictionaries, word counts, etc, and also by promoting/facilitating collaboration (Duriau, Reger, & Pfarrer (2007); Kabanoff (1997)).

An extensive research literature exists in the social sciences, which have made use of some form of content analysis. Recently, for example, Andreu, Caraganis, & Okazaki (2011) used content analysis to demonstrate how top-ranked MBA schools actively engage in SNS to provide rational, innovative and informative content to generate extensive networks of students and faculty members; Messinger (2012) described the utility of content analysis to social science research, and described an exercise using content analysis of Harry Potter to teach the principles of content analysis to undergraduate classes; Taylor, Francony, Beiting, Ritter, & Clutter (2011) demonstrated how content analysis was used to examine the composition of school wellness policies, to provide information about how the school environment influences children's dietary habits; Cukier, Trenholm, Carl, & Gekas (2011) used content analysis to study the emerging field of social entrepreneurship; and Rich (2012) used content analysis to demonstrate the use of automated data mining and content analysis of daily news reports from the Korean Central News Agency to track trends in North Korean rhetoric, and demonstrated that North Korea largely identified nuclear issues with the United States in contrast to other parties involved in the Six Party talks.

More relevant to the discussion of innovation, Folkestad & Gonzalez (2010) used content analysis to identify those organizational related themes that were associated with innovative work-places; Hienerth & Lettl (2011) used content analysis to look for factors that encourage innovation diffusion into mainstream markets; Rivas (2007) demonstrated the various types of organizational strategies that impacted firm performance in the ICT industry during technological disruptions; and Van Orden, van der Rhee, & Schmidt (2011) used content analysis to help identify how new products opened up new markets or encroached on incumbent products.

Of relevance to the current study, numerous authors have used the concept of attribute sets to investigate consumer preferences and discourse frames, etc. Attribute sets are collections of key synonymous words and ideas relating to a higher order concept, and can be used as the dimensions along which particular characteristics or features of a product or concept can be assessed against a need or a different concept. This is relevant, as to be disruptive an innovation must rank higher on dimensions influencing consumer preference than the incumbent technology, and often, general attitudes can be gleaned from identifying frames in discourse.

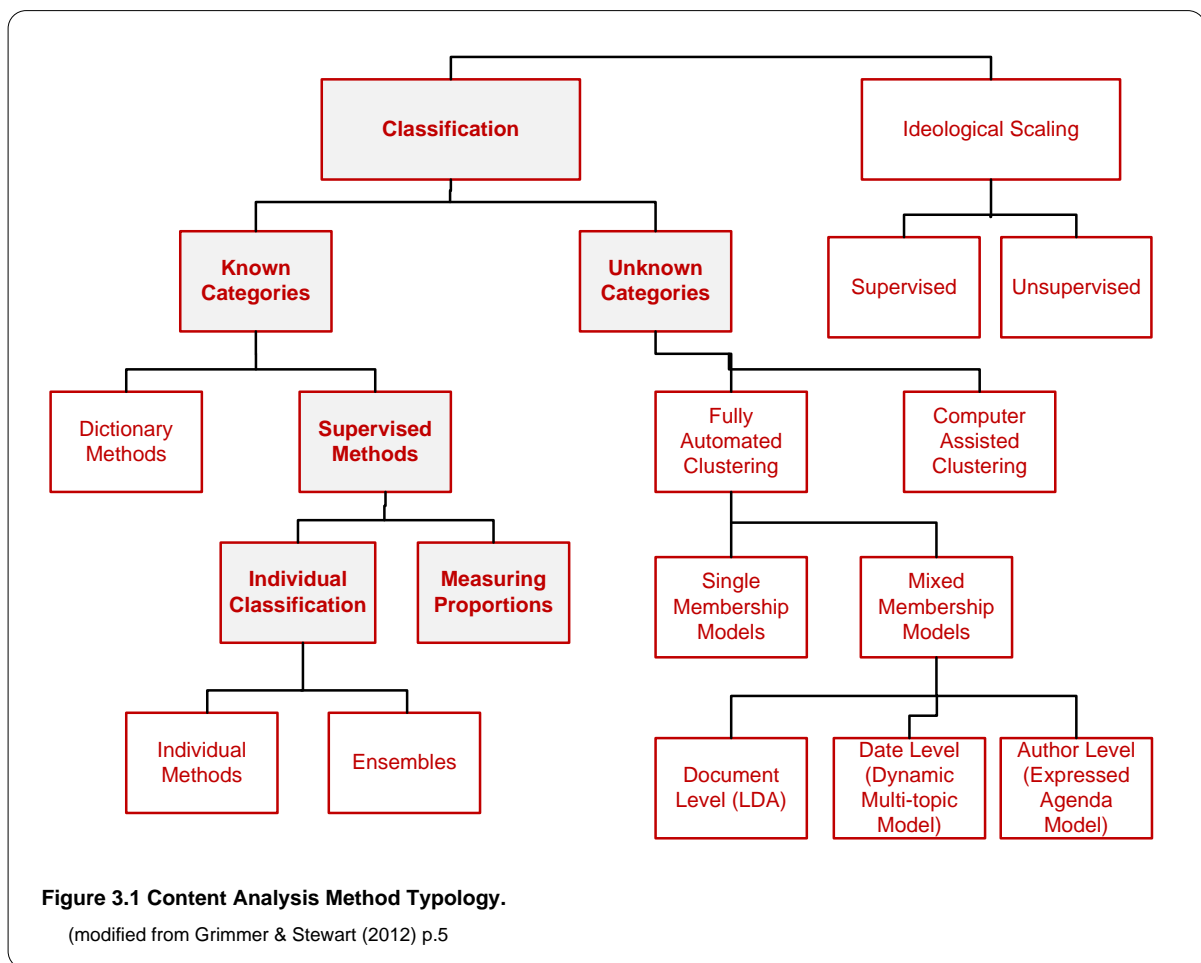
A number of examples can be found in the literature which have used attribute sets in this way (although not all were used in conjunction with content analysis). For example, Fiss & Hirsch, (2005) used the concept of attribute sets to identify frames within newspaper articles to study how the concept of globalization emerged and diffused; Rivard, Hautes, & Raymond (1997) used various attribute sets to assess software quality; and Usrey & Dooley (1996 & 1998) developed attribute sets to both define software quality and measure consumer attitudes towards it. More recently,

Owlia (2010) used attribute sets to define the quality dimensions used to assess quality of knowledge management systems, and Jiang, Shang, & Liu (2010) used attribute sets to define target items in a content-based filtering application of content analysis, in which they modeled and predicted consumer satisfaction.

In this study, attribute sets are used as a proxy for linguistic markers, which may be able to be used to identify disruptive innovations; content analysis is used as both an input into the creation of attribute sets, and also, in the determination of the frequency of occurrence of the attribute sets in the corpus studied. At the time of this writing, no reference for the use of attribute sets as linguistic markers could be found in the academic literature.

3.2 Content Analysis Process

There is a broad typology of content analysis methods. In a working paper, Grimmer & Stewart (2012) describe many of the methods used in content analysis, which view text as data. The following figure provides a visual overview of these methods. The methodology of this investigation falls under the broad classification branch of the typology, the various components touched on by the methodology used in this study are coloured light grey.



As can be seen in the figure 3.1, content analysis can be implemented in a considerable variety of ways. However, there are commonalities in methodological approach which transect them (Wolfe, Gephart, & Johnson, 1993), including the basic phases of data collection, coding, analysis of content, and interpretation of results (Duriau, Reger, & Pfarrer 2007).

The following sections briefly discuss the three basic phases referred to above, and also discuss the specifics of what was done in this study.

3.2.1 Data Collection

Data collection is typically the first phase of content analysis. In his textbook on content analysis, Weber (1990) identifies three key sampling decisions that need to be made, when it is not possible to utilize all (i.e. the entire population of) textual material:

1. Select the sources of information.
2. Define the type of documents that will form the basis of analysis.
3. Choose specific texts within the selected documents for analysis.

The outcome of these decisions will be driven by the research purpose, the methodology followed, and the availability of information.

Where content analysis hasn't been used prior to answer the research question, multiple sources of information are warranted (for the reasons outlined in Glola & Chittipedd (1991), who used content analysis in an ethnographic study of organizational sense making). Using a wide variety of data sources in studies looking at phenomena for the first time, benefit from such an approach, as it improves the validity of the approach, by incorporating the perspectives from multiple participants (Gephart, 1997). A related question that needs to be addressed when using a wide variety of data sources relates to exactly what texts should be used for the analysis, to allow a coherent view to emerge. Lissack (1998) was one of the first authors to extensively describe the approach of concept sampling in content analysis, to address this question. In a more recent paper, Elo & Kyngäs (2008) describe concept sampling as a mechanism used to develop an understanding of the meaning, intention, consequences and context of communication and to identify critical processes. Gephart (1993) describes how databases can be used to collate such concept samples from multiple sources.

Computer assisted content analysis has proved beneficial in amassing text from a wide variety of data sources for content analysis, by facilitating the access to (and simultaneous consideration of) a vastly larger corpus of documents than would be possible using manual content analysis (Duriau,

Reger, & Pfarrer 2007). Additionally, numerous computerized databases are available (e.g. LexisNexis, Proquest, Academic Universe) that can be used to source electronic textual articles that can be used for content analysis. Surprisingly, in their study of academic content analysis papers published between 1980 and 2005, Duriau, Reger, & Pfarrer (2007) found very few researchers made use of such databases. Equally surprising given the temporal nature of comparable corporate information over time, Duriau, Reger, & Pfarrer (2007) also found that only a third of the research papers reviewed had a longitudinal design, and so did not exploit the time series nature of the information collected.

This Study

This study has both cross-sectional and longitudinal design elements, and looks to identify linguistic markers for disruptive innovation between two related technologies (traditional BI and SaaS BI) both at a single point in time (using data aggregated from the corpus over the entire period from 2000 to 2012), and also over time (at the level of individual years). Due to the compressed timeline of the study, only data already in an electronic format was studied.

Limited content analysis based research has been identified in the academic literature to answer the specific research question of this study; the identification of linguistic markers indicative of a disruptive innovation. For example, Kostoff et al., (2004) suggests the use of content analysis in developing technology roadmaps and trying to determine what technologies may be disruptive, although the authors do not themselves perform such an analysis. Due to this, it was decided that multiple sources of information were warranted for this study.

Data Source & Selection

The primary database used was LexisNexis Academic Universe, using simple natural Boolean based search strategies to identify texts relating to traditional BI, and also SaaS BI. The focus of this study was on practitioner and industry discourse, which is reflected in the LexisNexis source selected; namely Industry News Publications, which provides access to full text English language articles. The Industry News Group file contains a wide variety of international industry and trade publications for over fifty industries. As at 30 September 2012, there were 2416 distinct named industry publications in the file.

The searching strategy looked to extract data from non-academic articles from 1985 to the present. Non-academic articles were chosen as it was assumed that as SaaS BI is a newer technology, if there is a 'lag' between practitioner (i.e. industry-based) literature and academic literature, there would be greater chance of finding reference to SaaS BI in practitioner based publications. This assumption was supported by the research of Z. Lee, Gosain, & Im (1999), who found evidence to suggest that

such a lag exists, and suggested that the delay could be due to the length of the academic peer review process and the lengthy data collection and analysis process that occurs in academic research. Gonzales, Udo, Bagchi, & Kirs (2011) showed similar results between academic business intelligence and data warehousing research, and also posited the influence of the peer review process as a driver of the lag. These authors also found evidence to support the notion that academic research in this area occurs within a closed community of peers, which may make research findings less relevant to, and disconnected from, practitioners. Evidence of a lag between academic and practitioner based literature was also provided by MacDonald & Richardson (2011), who demonstrated a lag between academic management accounting studies and management accounting practice.

Further support for the decision to use practitioner based literature over academic sources relates to the fact that the publications of industry analysts (such as Forester, AMR Research, Gartner, IDC, Yankee Group, etc) are viewed as a key source of information, and drivers of technology diffusion. Pollock & Williams (2011) brand organizational industry analysts as the "Knowledge Institutions of Information Technology", and cite their importance in shaping IT product markets. Swanson (2010) describes how such organizations can play a pivotal role in the diffusion of technological innovations (by publically supporting technologies), while in an earlier paper, Pollock (2009) describes how an unfair neutral or negative review by such industry analysts can hinder the diffusion of innovation. It is anticipated that the views of industry analysts relating to SaaS BI will have a large influence on whether the technology is widely adopted in the market. It should be noted however, that few industry analyst publications can actually be accessed via LexisNexis; reference to such publications was typically "second hand" and in independent publications.

As shown in figure 3.2 below, the data collection strategy used in this study looked to separate retrieved articles into 4 discrete groups for analysis: a group consisting of all traditional BI related articles; a group consisting of all SaaS BI related articles; and then a further subdivision of these groups into yearly tranches.

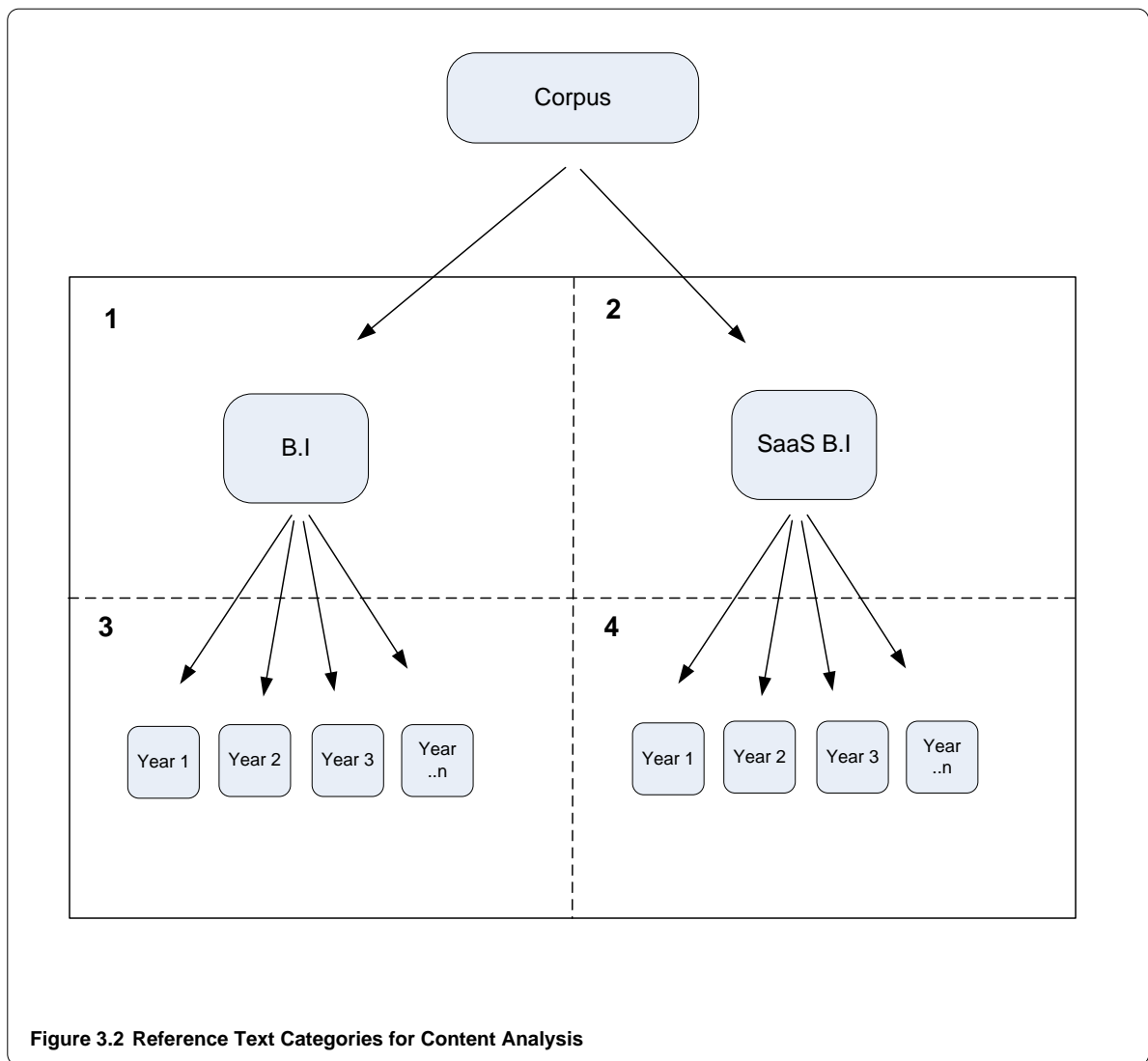


Figure 3.2 Reference Text Categories for Content Analysis

The purpose of this was to support a gross, time invariant analysis between BI and SaaS BI related articles, and also, a longitudinal analysis design, capable of detecting changes over time within each category. Analyses were performed on each of the groups.

The search string for traditional BI related industry news articles was :

"business intelligence" OR " bi " AND NOT ("saas business intelligence" OR "business intelligence saas" OR "bi saas" OR "saas bi" OR "software-as-a-service business intelligence" OR "software-as-a-service bi" OR "cloud based bi" OR "on demand bi" OR "on demand business intelligence" OR " bi-")

The search string for SaaS BI related industry news articles was:

"saas business intelligence" OR "business intelligence saas" OR "bi saas" OR "saas bi" OR "software-as-a-service business intelligence" OR "software-as-a-service bi" OR "cloud based bi" OR "on demand bi" OR "on demand business intelligence"

Note that LexisNexis searches are not case sensitive. Upper case words were only used in searches to improve readability. The returned output from searches is in the case provided by the publisher (invariably mixed case). The default font ("Courier") was used, to cater for fonts with varying font widths. Embedded graphics (e.g. photographs, charts and graphs, etc), were excluded.

A number of limitations within the LexisNexis software drove what date ranges (within the target date range) were used for each query:

- There is a limitation of only being able to identify search hits in blocks of 3000 or less (searches identifying more than 3000 hits result in a "more than 3000 Results" type message. Continuing with the query results in a truncated sample being returned)
- There is a 500 record limit for the number of records that can be downloaded at any one time
- The native download process is susceptible to network latency issues. Searches returning articles with multiple 'hits' against search terms often failed in the download process, as did full-text files with more than several hundred articles.

Date ranges were manipulated in response to the above limitations, on a search by search basis, to ensure that all articles fulfilling the search criteria could be identified and downloaded without issue and that the desired target date range was fully searched.

To overcome the difficulties in downloading files with full-text articles, the Key-Word-In-Context (KWIC) method was selected, where the LexisNexis search returns a 50 word block of text (from within the text of each article), 25 words either side of each occurrence of the key word being searched for. Once items were identified within LexisNexis database, they were manually downloaded from the LexisNexis web site, and saved as CSV files. Due to publisher restrictions, the LexisNexis retrieval process also includes publisher mandated information in addition to the KWIC contents. It was not always viable to remove this extraneous information. An assumption was made that the additional word content simply added "noise" to the overall result set, and that the component words of the extraneous information would be distributed randomly across the entire corpus of interest. This assumption was not tested however.

Simple DOS based file manipulation commands were then used to combine the individual downloaded files into single monthly files, which were then combined to form yearly files, as required.

3.2.2 Coding & Data Presentation

Coding relates to the identification and tagging of recording units of interest. Moretti et al., (2011) suggest that there are two main options when one considers broad approaches to coding; inductive (emergent) and deductive (a priori). In inductive content analysis, which is the conventional form of content analysis, coding categories are derived inductively from the raw data; that is, researchers take a viewpoint similar to that taken in Grounded Theory, and allow the data to drive theory development, so that the theory and analysis is grounded in the data (Punch, 2005, p.155).

Deductive content analysis tends to be more structured than the inductive approach (Moretti et al., 2011). The deductive approach works with prior formulated, theoretically derived categories (Mayring, 2000). The initial coding starts based upon theory or related research findings. Typically the deductive approach employs fixed word lists (dictionaries) in analyzing textual data (Mossholder, Settoon, Harris, & Armenakis (1995), Bengston, Potts, Fan, & Goetz (2005), Webb, Bengston, & Fan (2008)).

Indices are then constructed through counts/compilations of keyword scores on theoretically meaningful dimensions. Key to this approach is the assumption that word meaning is stable across groups or contexts, and that they reflect some a priori theoretical construct (Mossholder et al., 1995). The consistency and stability of individual word meaning over time is also key to the success of this approach (Slapin & Proksch (2008), (Martens, Poeppelbuss, & Teuteberg (2011)). Interestingly, Ronda-Pupo & Angel (2012) showed that while a 'nucleus' of key words can help define a concept over time, changes in the relative contribution (weighting) of individual words making up this nucleus over time, can alter the definition of the concept, regardless of the meaning stability of the individual words over the same time frame.

The idea that texts can be represented on different levels of abstraction, but that there are core meanings with insignificant variation, dominates dictionary approaches. In this approach, word families which can be used to identify common meaning between texts, form the basis of dictionaries. The dictionary approach can thus overcome the fact that language provides multifarious stylistic variations and expressive forms for the same concept (Krippendorff, 2012, pp.239-245).

Numerous authors have demonstrated the importance of the coding scheme to the success or otherwise of the content analysis exercise (e.g. Zhang & Wildemuth (2009); Soojung & Snaghee (2009); Vo (2012); Hopkins & King (2010)). The importance of coding is especially true for latent content analysis (Carley, 1993).

A coding protocol referenced often in the literature is Weber's Protocol (Neuendorf (2011); Tangpong (2011); Mantzoukas (2009); Hysong et al., (2009); Kirilenko & Stepchenkova (2012)). This protocol suggests that there are eight steps involved in the coding process that should be followed to minimize rater bias (Webber, 1990). The eight steps making up the Weber Protocol are listed below:

1. The definition of the recording units (e.g. word, phrase, sentence, paragraph, etc).
2. Definition of the coding categories (may be encapsulated in existing or self-created dictionaries, etc).
3. Test of coding on a sample of text.
4. Assessment of accuracy and reliability of sample coding.
5. Revision of coding rules.
6. Iterate through the process from step 3 until adequate reliability is achieved.
7. Code all of text.
8. Assess reliability or accuracy of output.

As in the data collection phase, computers can aid the coding phase, although a number of authors suggest that a method consisting of both automated coding with manual validation yields better results (e.g. C. H. Yu, Jannasch-pennell, & Digangi (2011), Sjovaag & Stavelin (2012)).

This Study

This study used a mix of both inductive and deductive content analysis. Basic content analysis, was undertaken using the Veneficium WordFrequencyCounter software (version 1.1), on a 64bit Windows Vista personal computer. The tool reads in CSV files, and produces frequency counts for the 100 highest frequency words in the file. It also provides the ability to use custom exclude/stop word lists, which removes unwanted words from the frequency counts. A stop word file was created from this study, based upon a version of a stop file sourced from the software tool QDA Miner (produced by Provalis research) and manually modified to remove additional unwanted words identified in the initial analysis phase of this study.

The Weber Protocol formed the basis of the coding method used in this project, as described below:

1. Recording units. The recording units for this study were the syntactical units of words and sentences, identified using LexisNexis. Due to the way LexisNexis performs KWIC searches, a 50 word block of text is retrieved for each occurrence of a key word, which often resulted in a text block with one full sentence with one or two additional partial sentences. A key assumption is that the literal and latent meaning of words over the time scale of this study (13 years) remains constant.

2. Coding categories were in the form of attribute sets, as described earlier. The ten categories for this study were driven by the literature search, and consisted of the categories Benefit, Consumer, External Environment, Feature, Industry Analyst, Industry, Output, Task, Technology, and Vendor. The categories are described in more detail later in this section.
3. A coding scheme was developed. In this instance, it simply consisted of attributing words from the lists of high frequency words identified using the software, as belonging to one of the 10 categories. A random selection of occurrences of the target word in the KWIC output text blocks was used to determine if the context of word usage impacted what category the word fell within. Running lists were maintained for each category during the coding process, to ensure consistency. Words that were identified as not belonging to any of the categories, or fell within multiple categories were discarded from the study.
4. For the cross sectional analysis looking across the entire data sets for traditional and SaaS BI data sets, a random sample of texts was selected for analysis by WordFrequencyCounter for the traditional BI articles, as more articles existed than could be processed by the software. The sample represented approximately 15% of the total corpus.
5. A manual assessment of the coding was made (for accuracy / validity). This was undertaken by sampling full text articles containing particular words categorized as belonging to the various attribute sets, to ensure correct inferences were being drawn. It must be noted that multiple raters were not used in determining the attribute sets, so inter-rater agreement was not able to be calculated. Due to this fact, it could be argued that the analysis is susceptible to the vagaries of subjectivity and interpretive bias, although as the categorization of words into attribute sets tended to rely on manifest meaning as opposed to latent meaning, the significance of this fact is diminished.

Attribute Set Categories.

Using the characteristics of disruptive innovations and the environmental conditions indicative of disruptive innovations as summarized in tables 2.1 and 2.2 (Section 2.1.9.1.2), a number of generic categories were derived to be used as the small class set identifiers for a number of attribute sets. The attribute sets were used in subsequent analysis, where comparisons between the frequency of their use in practitioner discourse relating to traditional BI and SaaS BI were made, to try and identify different usage patterns between traditional BI and SaaS BI. Such differences may result from the different stages of the product and technology lifecycle that each technology was in. Differences may indicate that SaaS BI is a disruptive technology, and ultimately form the basis for high level linguistic markers, for the ex ante identification of disruptive innovations. The following categories were used.

- *Benefit attributes*: This attribute set contained words indicating that a given technology was important to the technology consumer, and that they gained some value from it. Those technologies providing the best benefit to consumers were viewed as being more likely to capture market share at the expense of rival technologies.
- *Consumer attributes*. This attribute set consisted of words that related to the technology consuming organization, and was used as an indicator that vendors were targeting specific consumer functions, or aligning their product with specific market niches, etc.
- *External Environment attributes*: Words in this attribute set made reference to particular features of the external environment, such as markets, the size of the target/consuming organizations, etc. Such attributes may allow inferences to be made about the specific phase the market is in, etc.
- *Feature attributes*. This attribute set consisted of words relating to features of the technology that may be leveraged for some organizational benefit. Typically, disruptive technologies initially have features that do not compete directly with incumbent technologies, but over time, provide features which the market finds more desirable than those offered by the incumbent technology.
- *Industry Analyst attributes*. Words in this attribute set referred to industry analysts, either directly (e.g. Gartner), or indirectly via reference to an industry analyst publication or tool, or reference to research studies, etc. The comments of industry analysts can have a large bearing on the success or otherwise of a new technology or product.
- *Industry attributes*. This attribute set consisted of words relating to various industry verticals. The number of industries associated with a particular technology may give an indication on the phase that the market is in; whether the technology is mature, etc.
- *Output attributes*. Words relating to outputs of the consuming organization made up this attribute set. Often such words provide a tangible link between important organizational outputs (such as a report or product), and a particular supporting technology. Such links can result in inertia on the part of consumers to initially change to a different technology / product.
- *Task attributes*. This attribute set consisted of words relating to tasks /business processes undertaken by the consuming organization. The link between a product and business process may indicate the product is of strategic importance (to both the consumer and producer) etc.
- *Technology attributes*. The words in this attribute set relate to technologies within the consuming organization that the product interacts with or supports in some way. If the consumer technology is newer, it may infer that the vendor product is more novel; it may also infer that the product integrates with the consumer's technologies, etc.

- *Vendor attributes.* This attribute set consists of words that relate to vendors directly by name, or by reference to one of their products. Changes in the number and size of vendors in the market, can allow inferences to be made regarding the stage that the market is in.

The following table links the above categories of attribute sets to the various characteristics summarized in tables 2.1 and 2.2 that have been linked in the academic literature to disruptive innovations. The numbers are the ID of the particular characteristic, referenced in the 2 tables. It was possible for multiple attribute sets to relate to the same characteristic of disruptive innovation, and also for the same characteristic to relate to more than one attribute set.

Table 3.1. Category to Disruptive Innovation characteristic/supporting factor mapping.

Benefit: 1,2,3,4,7,8,11,13,22,34	Industry: 10, 19, 20
Consumer: 10, 34	Output: 5, 17
External Environment: 8, 15, 29, 34	Task: 16
Feature: 1,2,7,19	Technology: 5, 19
Industry Analyst: 1,2	Vendor: 10, 14, 18, 20, 23, 25, 29, 30

Note that it was not possible to address all the identified characteristics of disruptive innovations using the LexisNexis Industry Market Publications database. Table 3.2 on the following page lists those keywords from the KWIC text blocks that were categorized as falling within a particular category, and also provides some examples of the actual text blocks, with key words highlighted. Note that text blocks can contain key words spanning multiple attribute sets, meaning that a given text block can be coded as contributing to more than one category. In the table, this is shown by the same text block occurring more than once (associated with different categories). Only key words with a corpus frequency great enough to put them in the top 100 most frequent words are shown.

3.2.3 Analysis and Interpretation

The analysis and interpretation of stage of content analysis consists of two primary categories; quantitative and qualitative.

Quantitative Content Analysis

Quantitative analysis originally developed to study mass communication (Zhang & Wildemuth, 2009). Quantitative content analysis counts manifest textual elements, and typically involves provision of exact counts of word frequencies and semantic frequencies in those texts examined, and outputs words and phrases which can be categorized into semantic groupings. A criticism of

Table 3.2. Characteristics of Disruptive Innovations.

	Key Words	KWIC Text Block Example
Benefit	Ability, adaptable, advanced, affordable, agile, Apple, benefits, better, capabilities, collaboration, competitive, competencies, competency, complete, compliance, comprehensive, costs, critical, customer focused, cutting edge, dedicated, deliver, easy, e-business, e-commerce, effective, enterprise-grade, essential, experience, expertise, fast, financial, first, flexible, founder, global, growth, help, helping, improve, improving, improvement, increases, in-depth, insight, integration, iPad, key, latest, like, management, must, next-generation, novel, outcomes, performance, pipeline, powerful, proactive, progressive, prove, real-time, recovery, revenue, shares, simple, simply, simplify, specialist, specialized, standards, strategic, strategy, sustainable, targeted, unique, utility, value, valued, visibility, web, web-based, well, world, worldwide	<p>...The Business Intelligence competency demonstrates that RCG IT earns customer loyalty by helping leverage data to achieve more effective business results. RCG IT designs, deploys and customizes business intelligence solutions.</p> <p>...Being able to analyze business information to arrive at business intelligence is key to effective strategic management.</p> <p>... projects as a turn-key, comprehensive entertainment solutions provider, with global sales and marketing; physical and digital distribution, business intelligence tools, solutions and support ...</p> <p>...In its description of the appeal of SaaS BI, Gartner said, "Customers appreciate a quick time to value, lower upfront costs and the ability to leverage external expertise –</p> <p>... today announced the launch of its Conversion Action Platform, a Software-as-a-Service (SaaS) offering that combines web analytics and phone call tracking to simplify business intelligence and customer capture for SMBs.</p>
Consumer	Finance, marketing, operations, people, sales, small	<p>...which provides business intelligence software, is offering companies free online benchmarking. Finance executives can sign up for a scorecard that will ...</p> <p>... projects as a turn-key, comprehensive entertainment solutions provider, with global sales and marketing; physical and digital distribution, business intelligence tools, solutions and support ...</p>
External Environment	Companies, demand, market, markets, manufacturers, online, organizations, retailers, SMBs, website, websites	<p>... today announced the launch of its Conversion Action Platform, a Software-as-a-Service (SaaS) offering that combines web analytics and phone call tracking to simplify business intelligence and customer capture for SMBs.</p> <p>...acting in sync with reactive business intelligence, predictive RDI can give retailers the ability to act on shifting demand trends as they happen. RDI platforms are often the ...</p>
Feature	Agent-based, agents, application, applications, architectures, broader, collaborative, dashboard, on-demand, designed, distributed, e-mail, embedded, enterprise, ETL, e-trust, federated, hosted, infrastructure, internet, MDM, metadata, mining, mobile, on-demand, on-line, PC, PCs, portable, predictive, RDM, remote, SOA, statistical, subscription, suite, transaction	<p>...Acting in sync with reactive business intelligence, predictive RDI can give retailers the ability to act on shifting demand trends as they happen. RDI platforms are often the ...</p> <p>... mobile access to intranet and extranet portals running on SharePoint, which is becoming a corporate hub for data integration</p>
Industry Analyst	Gartner, Magic Quadrant, research, Reed, study	<p>...In its description of the appeal of SaaS BI, Gartner said, "Customers appreciate a quick time to value, lower upfront costs and the ability to leverage external expertise -</p>
Industry	Agriculture, agricultural, banking, biotech, drug, energy, entertainment, farmers, government, govt., health, healthcare, hospital, hoteliers, industry, insurance, manufacturing, medical, medicine, pharma, pharmaceutical, pharmacy, physician, retail, science, telecommunications, telecoms	<p>... projects as a turn-key, comprehensive entertainment solutions provider, with global sales and marketing; physical and digital distribution, business intelligence tools, solutions and support ...</p>
Output	Decisions, information, product, products, report, reports, reporting, solution, solutions	<p>... projects as a turn-key, comprehensive entertainment solutions provider, with global sales and marketing; physical and digital distribution, business intelligence tools, solutions and support ...</p> <p>...The Oracle Business Intelligence solution radically changes all our analytics and reporting processes. Its main advantage is instant data access and the speed of reports</p>

		creation, which ...
Task	Analysis, analytics, analyze, development, infomatics	<p>...The Oracle Business Intelligence solution radically changes all our analytics and reporting processes. Its main advantage is instant data access and the speed of reports creation, which ...</p> <p>...development including versioning, data consistency, database and business intelligence, tagging and rendering to outline the challenges ...</p> <p>...Being able to analyze business information to arrive at business intelligence is key to effective strategic management.</p> <p>... today announced the launch of its Conversion Action Platform, a Software-as-a-Service (SaaS) offering that combines web analytics and phone call tracking to simplify business intelligence and customer capture for SMBs.</p>
Technology	Chain, CRM, database, ERP, groupware, model, module, network, networking, portal, security, systems, warehouse, warehousing, wireless	<p>... mobile access to intranet and extranet portals running on SharePoint, which is becoming a corporate hub for data integration</p> <p>...development including versioning, data consistency, database and business intelligence, tagging and rendering to outline the challenges ...</p> <p>...times faster than other databases. Guess?, Inc. will run the Vertica Analytic Database in conjunction with front-end business intelligence tools from MicroStrategy and BusinessObjects.</p> <p>...Back-office components: Open Text Connectivity Solutions Group's Hummingbird business intelligence system that extracts data from computer-aided dispatch and records management systems ...</p>
Vendor Attributes	AFS, AgriCharts, Aim, AltoSoft, AMGAS, AMICAS, Analytix, AnalytixOnDemand, Anceta, Anthes, Apex, AppExchange, Aptify, AssetMetrix, Autometrics, AWS, Axxera, Birst, Blink, Brightstor, Brio, Business Objects, CDI, Celequest, CleverPath, CMP, Cigivar, Cognos, Crystal, CSN, Cutting Edge, Etology, Humedica, HummingBird, Htoerion, IBM, Infomatica, Innovatech, IRON, Islandia, Jaspersoft, Kluwer, Kognito, Laval, Lawson, Linthicum, Logic, LucidEra, MedAssets, Microsoft, Microstrategy, Muzeview, Oco, OnDemand, OpSource, Oracle, Pamerica, partnership, Pervasive Software, PivotLink, Quickbooks, Razorsight, Redshaw, RetroBox, RightNow, Sage, Salesforce, SAP, SAS, Savard, Sea Tab, Seagate, Starvest, Sterling, Tanguay, Travelclick, Unicenter, Utilisphere, Visiquate, Visual, WCI, Windows, Wolters, Xactly, Zensar	<p>...times faster than other databases. Guess?, Inc. will run the Vertica Analytic Database in conjunction with front-end business intelligence tools from MicroStrategy and BusinessObjects.</p> <p>...Back-office components: Open Text Connectivity Solutions Group's Hummingbird business intelligence system that extracts data from computer-aided dispatch and records management systems ...</p>

quantitative content analysis is that it can miss syntactic and semantic content embedded within the text (Weber, 1990). Typically, quantitative content analysis is deductive and is used to test hypotheses or research questions based on prior empirical research / observation. Usually, quantitative content analysis requires a probabilistic sample selection method (e.g. random sampling) to ensure the validity of any statistical inferences made. Often, the outputs from quantitative analysis are series of numbers amenable to manipulation from statistical techniques, ultimately appearing as a statistical significance of the occurrence of words or concepts. Additionally, because of the probabilistic nature of quantitative analysis, more advanced statistical techniques can be used for analysis and interpretation, such as factor and cluster analysis, multiple regression analysis, etc.

Qualitative Content Analysis

Qualitative analysis developed in the fields of anthropology, sociology and psychology, and is concerned with providing insight into the latent content (underlying meaning) of physical messages (Zhang & Wildemuth, 2009). Qualitative analysis tends to be inductive, and is grounded within the textual data itself. The method for selecting texts for qualitative content analysis is usually driven by the researcher. The outputs from qualitative analysis are usually descriptions and typologies (often with supporting quotes from within the text itself, illustrating the perspective of the author. Often, only non-probabilistic descriptive statistics are used to convey the results of qualitative study. Many authors suggest that robust content analyses use a mix of both quantitative and qualitative techniques (e.g. Weber, 1990; Zhang & Wildemuth, 2009).

This Study.

This was a preliminary study, with no similar studies being found in the academic literature. Due to this, this study had a quantitative focus. High frequency words from both the SaaS BI and traditional BI corpora were obtained from the KWIC listings obtained from the LexisNexis Industry News Publications database, using the content analysis software WordFrequencyCounter. These words were manually categorized as falling into one of 10 predetermined categories (attribute sets). Words unable to be categorized were discarded from the data set. The frequency of occurrence of each category was derived by summing the individual frequencies of the member words of each category. A check was made to see if there was any co-occurrence of member words in a sample of KWIC output text blocks. Where there were (such as in multi-part vendor names for example), only the occurrence of one of the words was counted.

Longitudinal (time series) and cross sectional (time invariant) analyses were then performed. Initially, data were plotted graphically to visually look for any patterns. Single factor ANOVA tests

were then performed to look for significant differences between the SaaS BI and traditional BI categories.

4 Results & Analysis

Analysis was undertaken on the on the SaaS BI corpus and the traditional BI corpus, and focused on the following four areas:

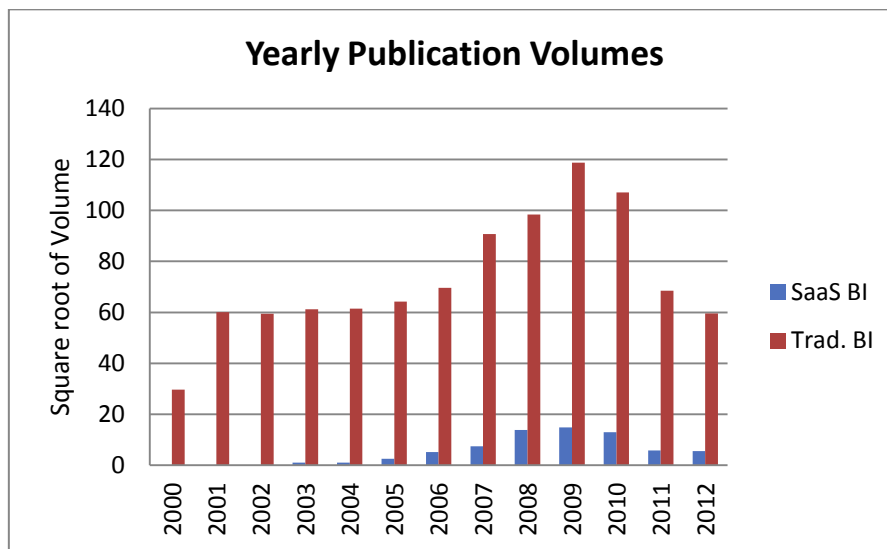
1. The article volumes for SaaS BI compared to traditional BI.
2. Analyzing the validity of using attribute sets in the current study.
3. Providing a longitudinal analysis of attribute frequency data for SaaS BI compared to traditional BI.
4. Analyzing the individual words from the corpora for SaaS BI and traditional BI (both shared across the corpora and unique to either corpus), to determine if any individual sets of words have the potential to be used as linguistic markers.

4.1 Article Volumes

Using the search strategies described in the previous section, a total of 76,213 articles were identified in the 13 year period from January 1st, 2000 through to 27 August, 2012, for traditional business intelligence. For the same time period, 732 articles were identified that related to SaaS BI. The disparity in article numbers is to be expected given the relative novelty of SaaS BI, and is consistent with Gartner's view that SaaS BI only has a market penetration of 1 to 5% (Gartner, 2011). Taking the lower end of the Gartner estimate (1%), a BI SaaS article count of 732 articles would correspond to a traditional BI article count of approximately 73,000.

To see if any discernible pattern existed in the volume of articles in the industry news for this time period between the two forms of BI, the volume data was transformed by using their square root, to help normalize the distribution, and make it easier to compare the volumes for traditional and SaaS BI on the same axes. Figure 4.1 plots these volumes.

Figure 4.1 Publication volumes between 2000 and 2012 for SaaS BI and Traditional BI.



As can be seen, from 2005, the distribution of the yearly volume of publications is similar for both groups, there being no obvious pattern that would enable SaaS BI to be distinguished from traditional BI, other than there being a much higher volume of articles for the traditional BI group. There is a decline in volume for both groups from 2009 to 2012; this is probably a reflection of the global financial crisis which took hold in 2008.

4.2 Attribute Set Stability

The highest frequency words for each year (for both SaaS BI and traditional BI) were categorized as belonging to one of 10 attribute categories. The attribute categories were generic but covered those areas that the literature search indicated may be useful in determining whether or not a technology innovation could be considered as potentially disruptive.

An assessment was made of the stability of the attributes sets used in the current study, on the grounds that stable attribute sets would be more suitable for identifying differences between the SaaS BI and traditional BI data sets. To do this, a cross sectional (i.e. time invariant) comparison was made between the traditional BI and SaaS BI groups, comparing unique words falling into the attribute categories for each group, and non-unique (i.e. shared) words falling into the attribute categories for each group, across the entire corpus. This assessment is shown in figures 4.2 and 4.3 below.

Figure 4.2. Cross sectional comparison of attribute sets composed of unique words in both the SaaS BI and Traditional BI data sets.

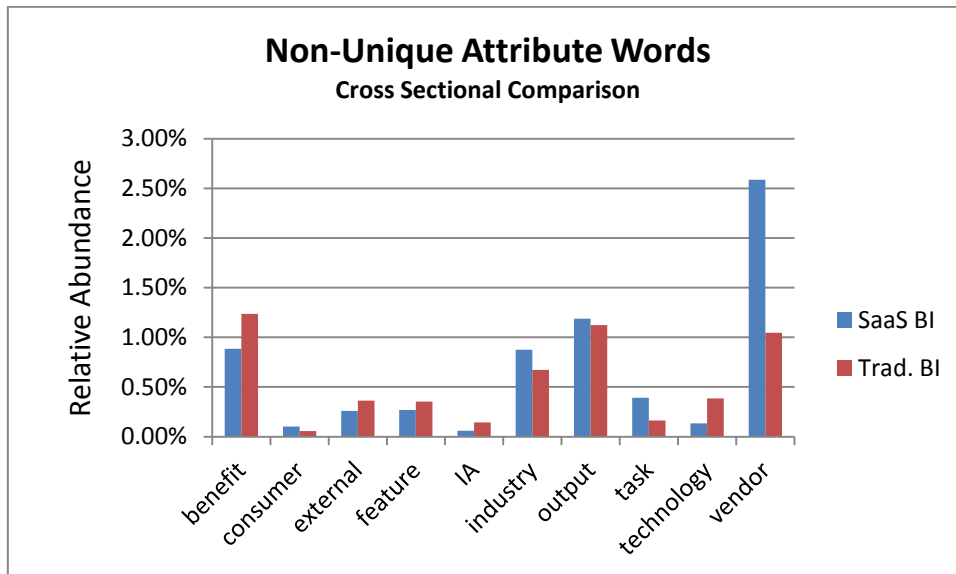
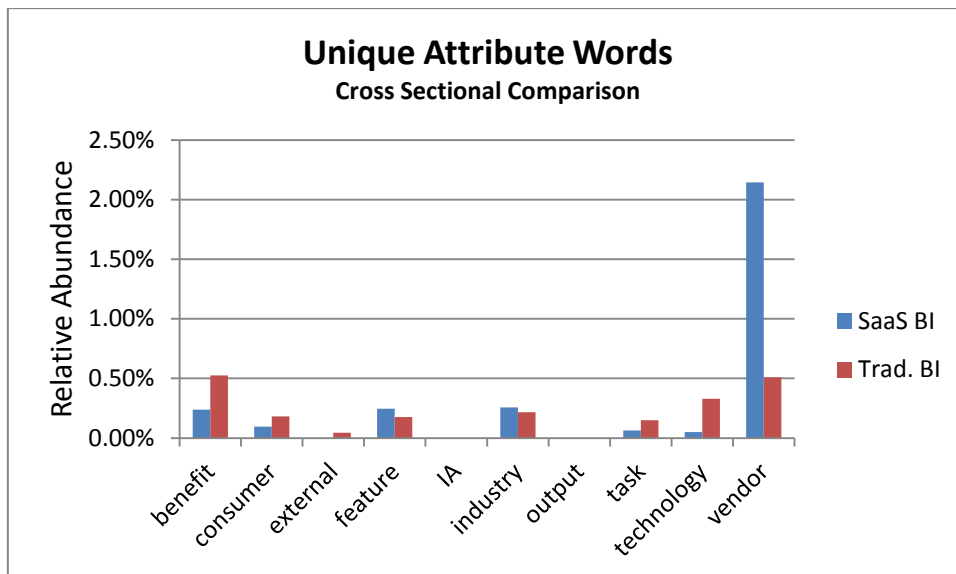


Figure 4.3. Cross sectional comparison of attribute sets composed of shared words in both the SaaS BI and Traditional BI data sets.



Figures 4.2 and 4.3 identify a number of consistent differences between SaaS BI and Traditional BI attribute sets, across both unique and non-unique words. For example, there are consistent differences between the benefit, industry, technology, and vendor attribute sets. The consistency between datasets composed of unique words and shared words across both SaaS BI and traditional BI may indicate that these attribute sets are stable enough at the level of latent meaning to be useful in distinguishing between SaaS BI and traditional BI.

4.3 Longitudinal Analysis of Attribute Sets

Analysis was undertaken to provide a longitudinal view of the occurrence of attribute categories within the corpus, to compare SaaS BI and traditional BI (on a year by year basis). The occurrence represents the occurrence of words belonging to an attribute set, expressed as a percentage of the total word count across all articles. The occurrence figures present a complex picture, as shown in figures 4.4 and 4.5 below. While the benefit attribute category seems to be rising for traditional BI and falling for SaaS BI, there are too few data points to make a definitive case at this stage, and no obvious patterns emerged. The peak values for the initial years for SaaS BI may be an artifact of the relatively low number of articles in the earlier part of the time series data, which may have resulted in some words being over represented.

Figure 4.4. Attribute category abundance across the corpus for Traditional BI.

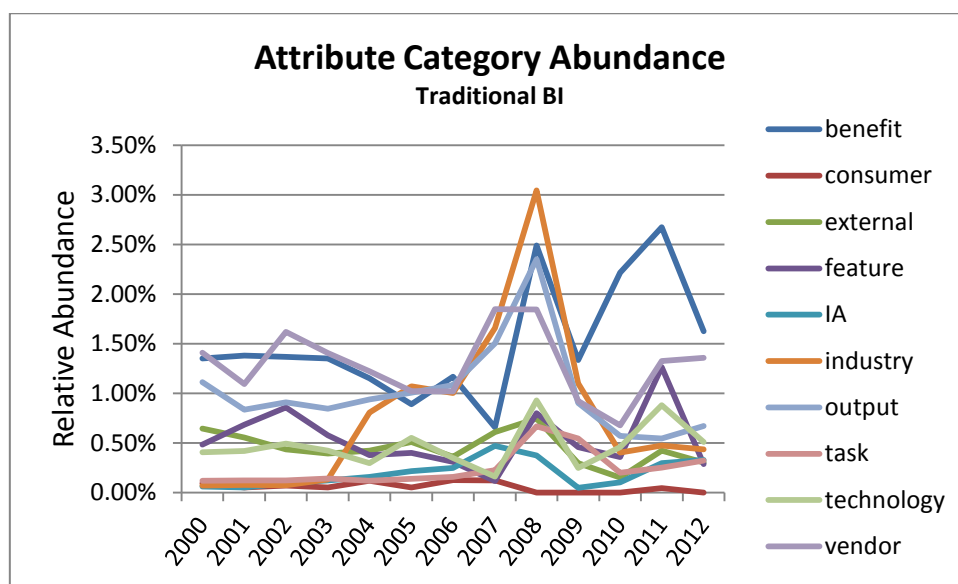
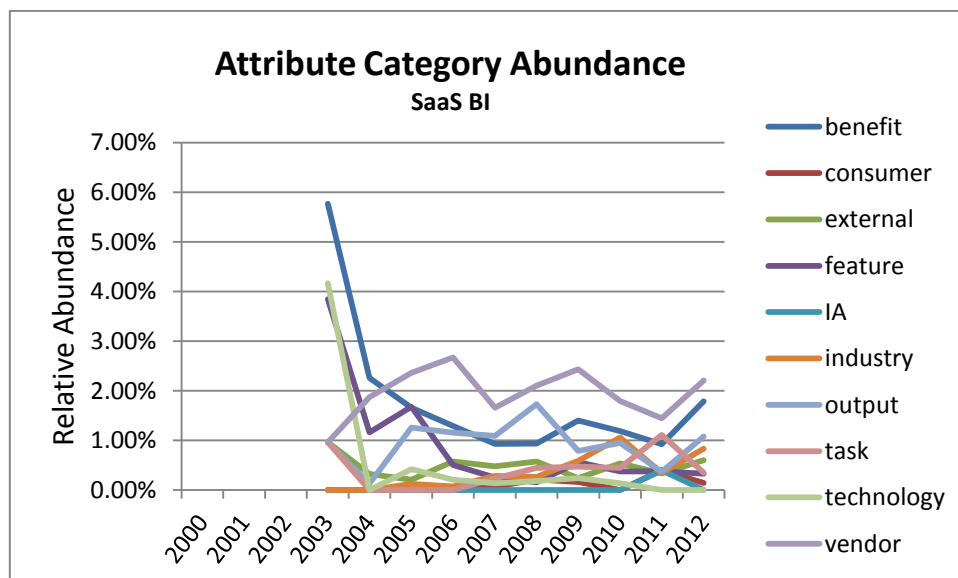


Figure 4.5. Attribute category abundance across the corpus for SaaS BI.



In an attempt to reduce the variation, and to see if any higher level patterns emerged, time series data consisting of absolute abundance values were examined. Relative abundance values were calculated against mean absolute abundance values (across all attribute sets, over a year). This measure indicates the relative movement of abundance values for each attribute category against all other categories for a given year. To further reduce complexity, the percentage of observations (for each attribute set) greater than the mean, less than the mean, and that were zero, were also computed. These values are shown in the figures 4.6 and 4.7 below.

Figure 4.6. Relative abundance of attribute sets for traditional BI

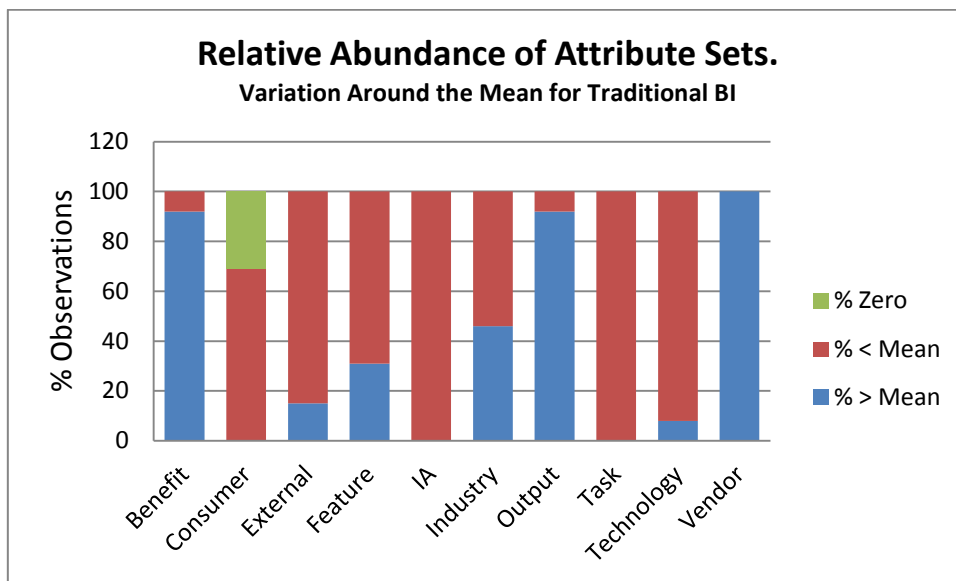


Figure 4.7. Relative abundance of attribute sets for SaaS BI

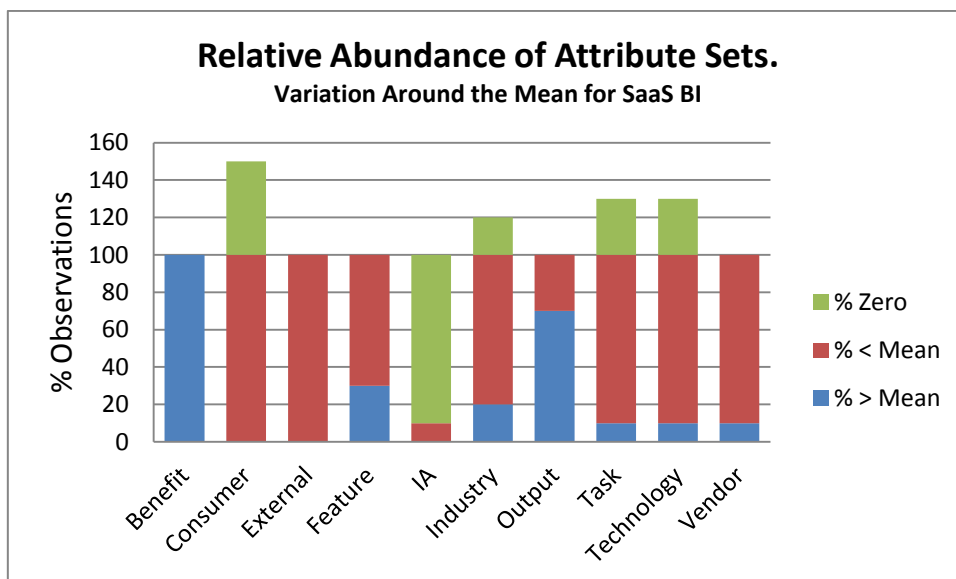


Figure 4.7 above shows for example, for SaaS BI, that the benefit attribute set had 100 % of observations greater than the mean, while 100% of the non-zero observations for the consumer attribute set were less than the mean, but 50 percent of the observations were zero. For most attribute sets, the figures demonstrate that there is a degree of stability when attributes are compared against each other over time (as indicated by there being few attribute sets that demonstrated variation about the mean approximating 50% of observations above the mean, and 50% of observations below the mean). This supports the notion that attributes sets are an appropriate mechanism to look for linguistic markers over time, and it may be possible to detect significant departures in relative abundance, when attribute sets are measured relative to each other (using appropriate statistical techniques).

Single factor analysis of variance (ANOVA) tests were performed on the attribute set abundance datasets for SaaS BI and traditional BI (where the abundance of attribute sets in the corpus was expressed as a percentage of the total word count across all articles). Table 4.1 summarizes the outcome for the ANOVA tests.

The test indicated that almost all of the attribute sets were significantly different when SaaS BI was compared to traditional BI. However, when the plots of the raw abundance figures are inspected (figures 4.4. and 4.5), no obvious patterns can be seen in the time series data.

Table 4.1. Significant differences between attribute set relative frequency

Attribute Set	Significantly Different	F ratio	F ratio : F crit	P value
Benefit	Y	18.93590	4.3	0.004
Consumer	N	4.47990	1.0	0.0485
External Env.	Y	24.97680	5.7	0.0001
Feature	Y	18.11482	4.1	0.0005
Industry Analyst	N/A			
Industry	Y	1.44652	2.6	0.0033
Output	Y	19.05238	4.3	0.0004
Task	Y	5.75920	1.3	0.0274
Technology	Y	43.45926	9.8	0.000003
Vendor	Y	25.358889	5.7	0.0001

It is assumed that this result has arisen because the abundance values are so different between SaaS BI and traditional BI, due to the much larger corpus for traditional BI; the volume of traditional BI articles available in the corpus was several orders of magnitude greater than was available for SaaS BI.

To try to overcome this, raw abundance values were expressed as relative abundance values, by expressing abundance against the mean absolute abundance values, across all attribute sets, over a year. This measure indicates the relative movement of abundance values for each attribute category against all other categories for a given year. Again, analysis of variance tests (single factor ANOVA) between the relative abundance data for the attribute sets for SaaS BI and traditional BI were run. Table 4.2 summarizes the outcome of these tests.

Table 4.2. Significant differences between attribute set relative frequency

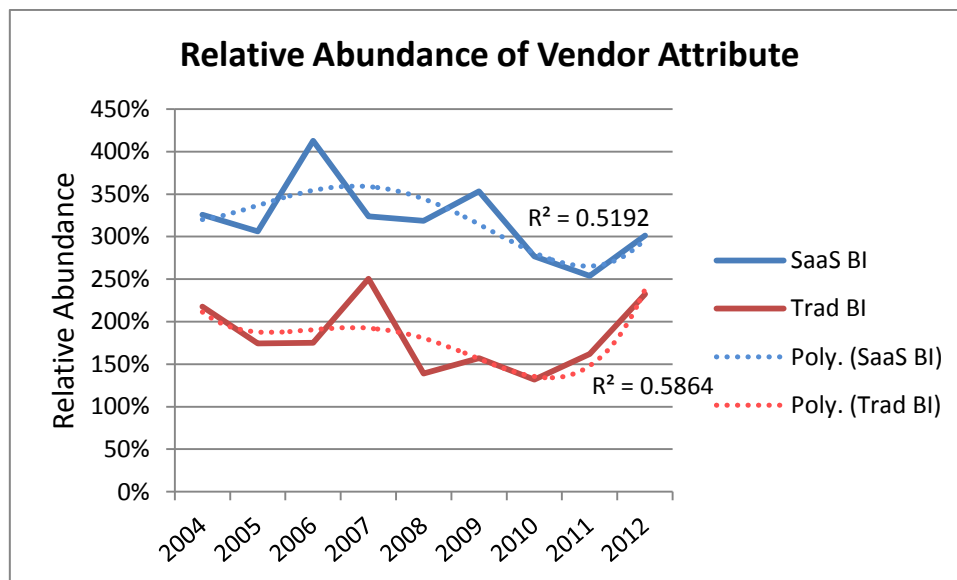
Attribute Set	Significantly Different	F ratio	F ratio : F crit	P value
Benefit	N	0.0611	0.01	0.80755
Consumer	N	0.93693	0.21	0.34590
External Env.	N	0.25102	0.06	0.62243
Feature	N	1.37930	0.31	0.25552
Industry Analyst	N/A			
Industry	Y	8.31334	1.88	0.00990
Output	N	0.28165	0.06	0.60211
Task	N	0.5929	0.13	0.45128
Technology	N	1.31389	0.30	0.26670
Vendor	Y	9.63523	2.18	0.00612

The ANOVA tests indicate that the industry and vendor attribute sets are different at a statistically significant level when SaaS BI is compared to traditional BI. This is consistent with a visual comparison of figures 4.2 and 4.3, and also 4.6 and 4.7. Simple visual inspection highlighted other potential differences, although the ANOVA tests indicate that the differences were not statistically significant. Both the Vendor and Industry attribute sets were investigated further to see if further insight could be gained into the nature of the differences observed.

4.3.1 Vendor Attribute Set Differences

One consistent observation is that the corpus makes significantly higher reference to vendor attributes for SaaS BI than for Traditional BI. Figure 4.8 below plots relative abundance figures (relative to the mean attribute abundance across all attribute sets for a given year) for the Vendor attribute. The years from 2004 through 2012 were used, as consistent data for both groups was available for this time series. A trend line is also plotted for each group; a 4th order polynomial was used, as it provided the best R² value.

Figure 4.8. Relative abundance of the Vendor attribute set for SaaS BI and traditional BI, for the years 2004 through 2012.

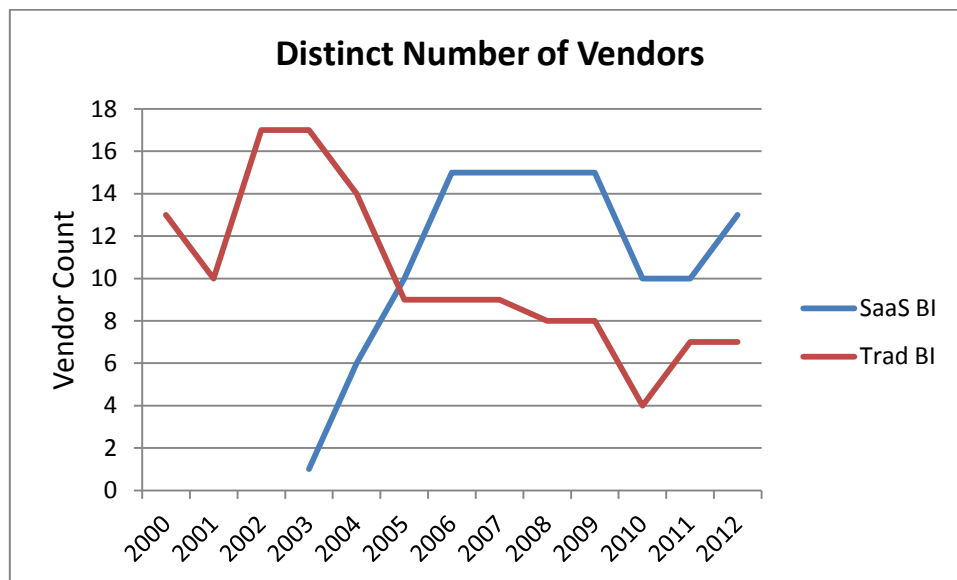


The plots for both groups of data are similar, and the trend lines follow a similar pattern that appears to reflect fluctuations in the yearly volume of the corpus. The volume of vendor references is consistently greater for SaaS BI however. A single factor ANOVA test was performed on this data (2004 through 2012). It indicated a statistically significant difference between the relative frequency of the Vendor attributes set for the SaaS BI group, and the relative frequency of the Vendor attributes set for the traditional BI group; the *F* value was 44.12, the *F crit* value was 4.49, and the *P-value* was < 0.0005).

The number of distinct vendors was also investigated to see if any obvious trend emerged. The distinct number of vendors represents a unique count of vendors for each group (SaaS BI and traditional BI), for each year, where the vendor name was frequently enough mentioned in the corpus to be in the 100 most frequent words selected by the analysis software.

Looking at the number of distinct vendors in each group over the period, it is seen that the trend for traditional BI is of declining vendor numbers, while for the SaaS BI group, the trend is for vendor number to increase. This is seen in figure 4.9, where the plots cross over in 2005.

Figure 4.9. Distinct vendor numbers for SaaS BI and traditional BI, for the years 2000 through 2012.



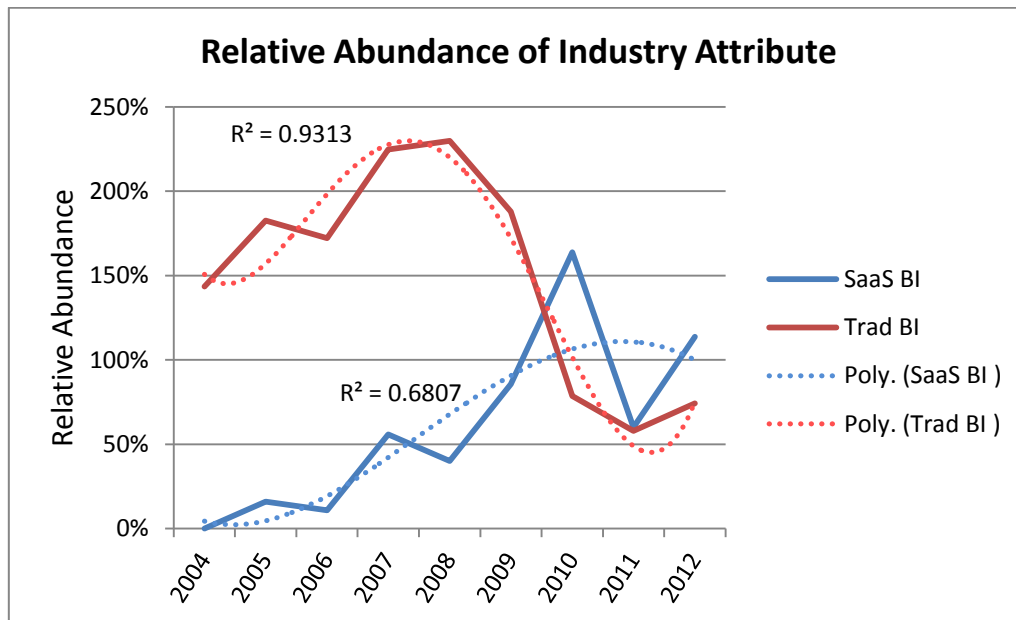
This observation is consistent with the SaaS BI market being in the fluid phase of the Abernathy-Utterback model (where there are many vendors), while for traditional BI, there are fewer vendors (consistent with the market for traditional BI being in the specific phase). When a single factor ANOVA test was performed on this data (2004 through 2012), there is a statistically significant difference between the number of distinct vendors in the SaaS BI group, and the number of distinct vendors in the traditional BI group; the F value was 12.57, the F_{crit} value was 3.38, and the P -value was 0.016. The sharp downward discontinuity in vendor numbers seen to occur in 2009 may again, be a reflection of the global financial crisis.

4.3.2 Industry Attribute Set Differences

The industry attribute set was the other attribute set that was shown to be statistically different between SaaS BI and traditional BI. The picture is a lot less clear for this attribute set than it was for the vendor attribute set however. Figure 4.10 plots relative abundance figures (relative to the mean attribute abundance across all attribute sets for a given year) for the Industry attribute. A trend line is also plotted for each group; a 4th order polynomial was shown to give the best R^2 value.

A single factor ANOVA test was performed on this data for years 2004 through 2012. It indicated a statistically significant difference between the relative frequency of the Industry attribute set for the SaaS BI group, and the relative frequency of the Industry attribute set for the traditional BI group; the F value was 8.31, the F_{crit} value was 4.41, and the P -value was 0.01).

Figure 4.10. Relative abundance of the Industry attribute set for SaaS BI and traditional BI, for the years 2004 through 2012.



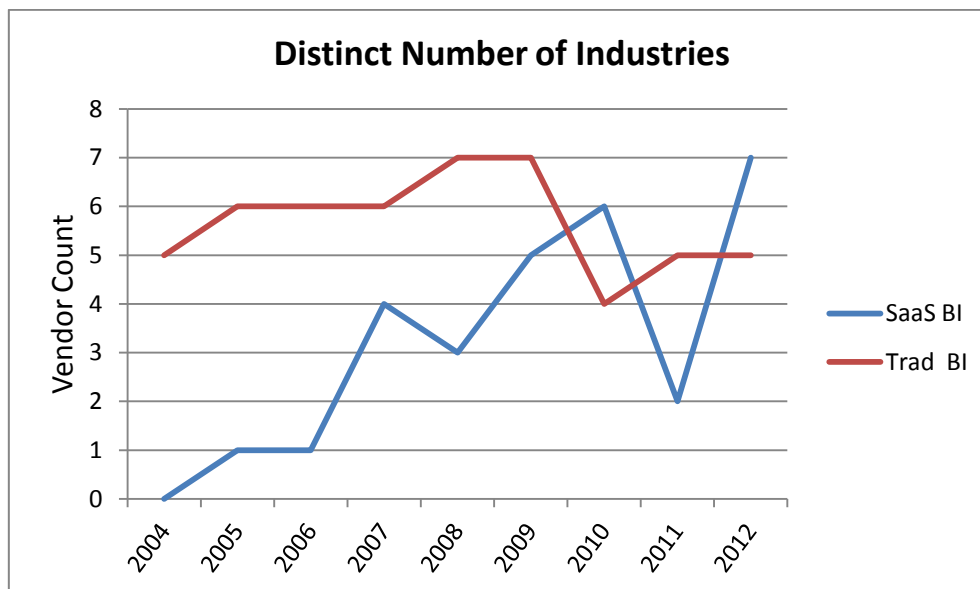
In figure 4.10, the relative abundance of references to industry attributes is seen to reduce in the corpus for traditional BI, while for SaaS BI it is seen to increase. This may indicate that the overall market is beginning to contract for traditional BI products, while at the same time, it is increasing for SaaS BI based products. This observation may be suggestive of the beginnings of a market displacement / disruption.

The number of distinct industries occurring in the corpus was also investigated to see if any additional insight could be gained. Figure 4.11 plots the number of distinct industries each year for both SaaS BI and traditional BI. When a single factor ANOVA test was performed on this data (2004 through 2012), there is a statistically significant difference between the number of distinct industries in the SaaS BI group, and the number of distinct industries in the traditional BI group; the F value was 7.74, the F_{crit} value was 4.49, and the P -value was 0.013).

The number of distinct industries seems to follow the trend for relative abundance of industry attribute sets, with the data for traditional BI trending downward over time, while for SaaS BI, the

data trend upwards over time. This could be consistent with the market for traditional BI being in decline, while for SaaS BI, it is in a phase of expansion.

Figure 4.11. Distinct industry numbers for SaaS BI and traditional BI, for the years 2004 to 2012.



4.4 Potential Linguistic Markers

It was assumed that linguistic marker words would show stronger association with either SaaS BI or traditional BI. The degree of association of individual words to SaaS BI and traditional BI can range anywhere on the spectrum, from the word being unique to one corpus through to being shared evenly between both corpora. Analysis was performed to determine words unique to each corpus, and then to investigate words shared across the corpora and the relative strength of their association to SaaS BI and traditional BI.

4.4.1 Unique words

A total of 29 words were unique to the SaaS BI Corpus, and 33 words were unique to the traditional BI corpus, when viewed as a whole across all time periods. When viewed at the level of attribute sets, the set of unique words in each corpus showed different distributions, as illustrated in figure 4.12 below. In the figure, attribute sets are shown a consistent colour between each donut chart, and the percentages cited, are the abundance figures as a proportion of the total word count for the data set. The figures support the view that there are potentially differences between the SaaS BI

corpus and the traditional BI corpus that are of sufficient magnitude to allow the two to be differentiated.

Figure 4.12. Relative frequency of attribute sets composed of unique words for SaaS BI and traditional BI.

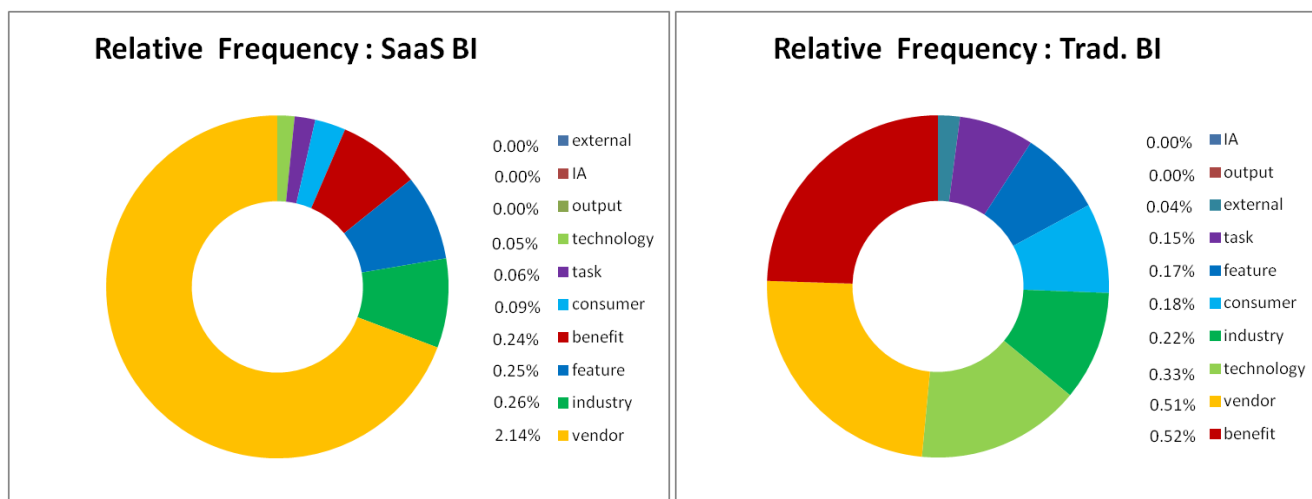


Table 4.3 on the following page lists the 25 highest ranking unique words from each corpus. The unique words for the SaaS BI corpus were mainly centered on vendors, industry verticals, product features, and benefits of the technology. The unique words for the traditional BI corpus tended to center around benefits of the technology, vendors, technological features of the consuming organization, industry verticals, and attributes of the consuming organization. It should be noted that the words are unique only within the context that the top 100 most frequent words made up the samples from which the unique words were drawn.

Table 4.3 Highest 25 ranked words from the SaaS BI and traditional BI corpora.

SaaS BI		Traditional BI	
Word	Relative Freq.	Word	Relative Freq.
Birst	0.347	products	0.138
Blink	0.254	report	0.112
Oco	0.224	SAS	0.100
AMICAS	0.196	healthcare	0.100
Pivotlink	0.194	financial	0.085
leading	0.193	real-time	0.084
Humedica	0.147	Oracle	0.077
demand	0.144	Edge	0.077
Sea Tab	0.108	analysis	0.075

BusinessObjects	0.106	development	0.074
Analytix	0.094	pharmaceutical	0.073
retail	0.082	systems	0.066
OnDemand	0.074	Microsoft	0.064
benefits	0.072	CRM	0.064
medical	0.071	capabilities	0.063
informatics	0.064	suite	0.062
next-generation	0.062	database	0.059
LucidEra	0.060	MicroStrategy	0.059
entertainment	0.056	strategic	0.058
value	0.056	key	0.057
mobile	0.051	network	0.057
Pervasive	0.051	improve	0.052
broader	0.051	IBM	0.047
model	0.049	Lawson	0.044
insight	0.049	operations	0.044

The trends visible in the words from each corpus are consistent with what could be expected when a comparison is made between an established technology, and a novel new technology. For example, 10 (40%) of the most frequent unique words in the SaaS BI corpus relate to vendors, only 1 of which is a globally recognizable brand (Business Objects). The traditional BI corpus by comparison had 7 (28%) of the most frequent words relating to vendors, and had 4 globally recognizable brands (IBM, Microsoft, Oracle, SAS). Apart from Business Objects (founded in 1990, although SAP, its parent company since 2008, was founded in 1972), the other vendors featuring heavily in the in SaaS BI corpus typically do not have the longevity of the vendors that feature heavily in the traditional BI corpus; for example, IBM was founded in 1911, Microsoft was founded in 1975, Oracle was founded in 1977, and SAS was founded in 1966.

This difference in vendor makeup will have contributed to some of the other obvious trends in the data. For example, the functions and features promoted in the SaaS BI corpus ('informatics', 'mobile', 'model', 'insight') are typical of words relating to newer technologies, while the functions and features promoted in the traditional BI corpus ('products', 'report', 'analysis', 'development', 'CRM', 'database', 'network') are typical of words that are associated established business intelligence technologies.

The same is true for the benefits promulgated in each corpus. For example, 'leading', 'next-generation', and 'on-demand' were high frequency words in the SaaS BI corpus, and are consistent with words relating to newer technologies, while 'real-time', 'strategic', 'improve', 'capabilities' from the traditional BI corpus, are words that you would typically associated with established technologies.

Similarly, both groups display differences in the industries that feature highly. For example, 'entertainment', 'mobile', and 'medical' from the SaaS BI corpus are all industries which have only recently begun to feature in popular consumer oriented ICT business literature, while 'healthcare', 'financial', 'operations', and 'pharmaceutical' from the traditional BI corpus, are all industries that have featured in popular ICT business literature for some time. It does not seem unreasonable to assume that these established industries would be served by the established BI vendors, while the more recent industries featuring in the SaaS BI corpus afford a niche market opportunity for the less well established SaaS BI vendors. Such a view is consistent with SaaS BI being a disruptive technology which is entering the market as a "low-end" disruption.

4.4.2 Words common across both corpora

The 36 most common words shared between the SaaS BI and traditional BI corpora were investigated to see if any specific words in the corpora had the potential to be used as linguistic markers. The most frequent words from each group across all time intervals were analyzed (for the traditional BI group, this necessitated using a random sample of 15% of the articles. The whole corpus was used for SaaS BI).

Words common to both the SaaS BI and traditional BI corpora were compared based upon their relative frequency against the total word count in their respective corpus. If there was no difference in the relative strength of a given word's association with SaaS BI or traditional BI, it would be expected that the relative frequency values for the word across both groups would be similar.

A Chi square (χ^2) test was used to provide an indication as to whether the null hypothesis (H_0) that any differences in word frequency observed between the SaaS BI and traditional BI groups can be accounted for by assuming the difference is due to chance. A significance value of 0.05 was initially selected as the value below which the hypothesis would be rejected.

Table 4.4 lists the 36 most common shared words, provides relative frequency values for both corpora, and the calculated Chi square value. As can be seen, there are 6 potential candidate words

that could comprise a linguistic marker enabling SaaS BI to be distinguished from traditional BI, or 9 words if the significance threshold is raised to 0.06. In the table, words with a Chi Square significance value of 0.06 or below are highlighted. The words are analytics, enterprise, global, information, management, partner, research, SAP, and solution.

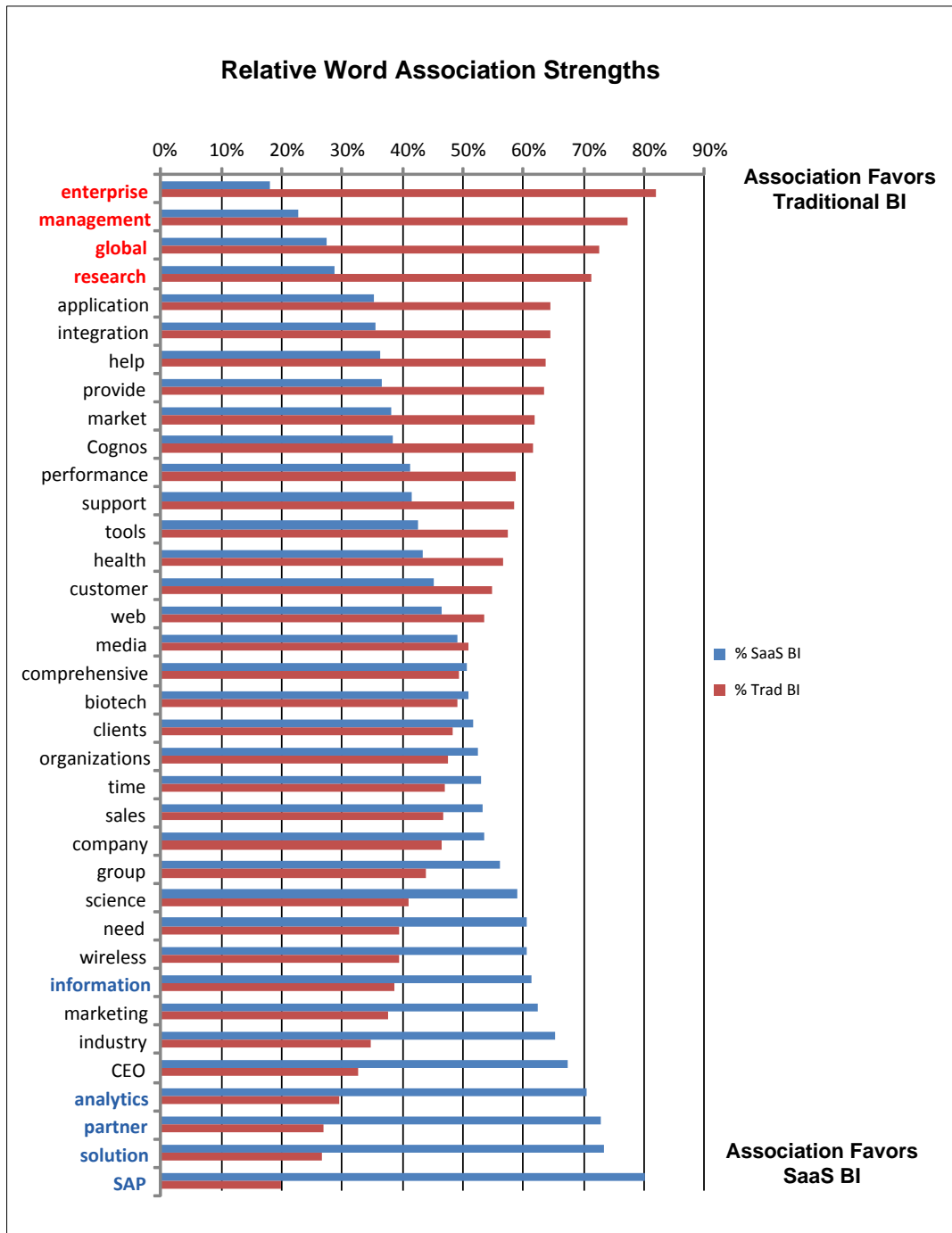
Table 4.4 Highest 36 ranked shared words from the SaaS BI and traditional BI corpora.

Word	Freq. SaaS BI	Freq. Traditional BI	Chi Squared
analytics	39.01	16.41	0.002
application	7.50	13.68	0.179
biotech	5.55	5.33	0.946
CEO	8.89	4.30	0.207
clients	5.97	5.57	0.907
Cognos	5.83	9.32	0.370
company	30.27	26.26	0.595
comprehensive	5.28	5.17	0.973
customer	7.77	9.39	0.696
enterprise	5.28	24.02	0.001
global	5.69	15.06	0.040
group	8.89	6.98	0.632
health	19.16	24.91	0.386
help	5.00	8.79	0.307
industry	17.08	9.07	0.117
information	44.57	28.00	0.052
integration	6.25	11.39	0.221
management	7.91	26.95	0.001
market	11.25	18.29	0.195
marketing	10.27	6.19	0.314
media	20.13	20.97	0.895
need	6.94	4.51	0.472
organizations	5.41	4.92	0.877
partner	12.91	4.79	0.054
performance	7.64	10.86	0.453
provide	5.41	9.42	0.298

Word	Freq. SaaS BI	Freq. Traditional BI	Chi Squared
research	5.97	14.80	0.053
sales	8.33	7.34	0.803
SAP	16.94	4.22	0.006
science	8.75	6.10	0.493
solution	66.78	24.42	0.000009
support	5.28	7.43	0.546
time	4.72	4.17	0.852
tools	7.91	10.62	0.529
web	5.83	6.70	0.807
wireless	8.47	5.49	0.425

To aid visual inspection of the association data, a tag cloud was created to provide a high level overview (figure 4.13). A more detailed view is then provided in chart form, in which the relative frequencies for each word were summed and the contribution of each proportion (SaaS BI and traditional BI) was expressed as a percentage of the sum, and plotted as a bar chart. This is shown in figure 4.14. In this figure, those words indicated by the chi square test as potentially having a stronger association towards either SaaS BI or traditional BI not explainable on the basis of chance alone are highlighted in red or blue, depending upon whether the association favored traditional BI (red) or SaaS BI (blue). The significance level was set at 0.06. The relative strength of the association is indicated by the two bars for each word, and the words have been ordered so that the association favors traditional BI at the top of the Y axis and SaaS BI at the bottom. The word 'analytics' for example, has a stronger association with SaaS BI, to the extent of approximately 2.3 to 1.

Figure 4.14. Relative word association to both SaaS BI and traditional BI.



It should be noted that 'CEO', 'industry', and 'marketing' (all words not considered significantly different between the SaaS BI and traditional BI corpora at the 0.06 level) sort between 'information' and 'analytics', due to their lower relative frequency values. For example, across both corpora the average relative frequencies were 6.6 for 'CEO', 13.07 for 'industry', and 8.23 for 'marketing'. For 'information', the average relative frequency was 36.92 and for 'analytics' it was 27.7. The chi square test is less sensitive to differences between numbers which have lower expected values, especially for values between 1 to 10 (Spiegel, 1992, p. 247).

A number of observations can be made regarding figure 4.14. Of the words found to be significantly different between the two corpora, 4 were more strongly associated with traditional BI. They were 'enterprise', 'global', 'management', and 'research'. These words are all words one would typically associate with established BI technologies and vendors; established BI technologies are enterprise level tools, often used by multinational organizations to support both local and global decision making.

It is not surprising that the word 'enterprise' is more strongly associated with traditional BI than SaaS BI, as small to medium businesses (SMBs) are the group most expected to adopt SaaS BI technologies (Sheikh, 2011), and will be the group targeted most by vendors - so words such as enterprise could be expected feature more prominently in the traditional BI corpus. Similarly, management could also be expected to be more prominently associated with traditional BI, due to traditional BI often being associated in the organizational setting with strategic management, operations management, risk management, performance management and balanced scorecards (e.g. Elbashir, Collier, & Sutton (2011), Kumar (2012)).

Five words were found to statistically be more strongly associated with SaaS BI. They were 'analysis', 'information', 'partner', 'SAP', and 'solution'. As for the unique words that associated more strongly with SaaS BI, these words could be seen as being more pertinent to newer technologies. For example, links could be drawn between the words 'analysis' and 'information' with the rising concept of "Big Data", a topic of contemporary interest in the business analytics field (Lavalle & Lesser et al., 2011). They would also appear to be congruent with the newer concepts of Data-as-a-Service and Analysis-as-a-Service (Stipic & Bronzin, 2012).

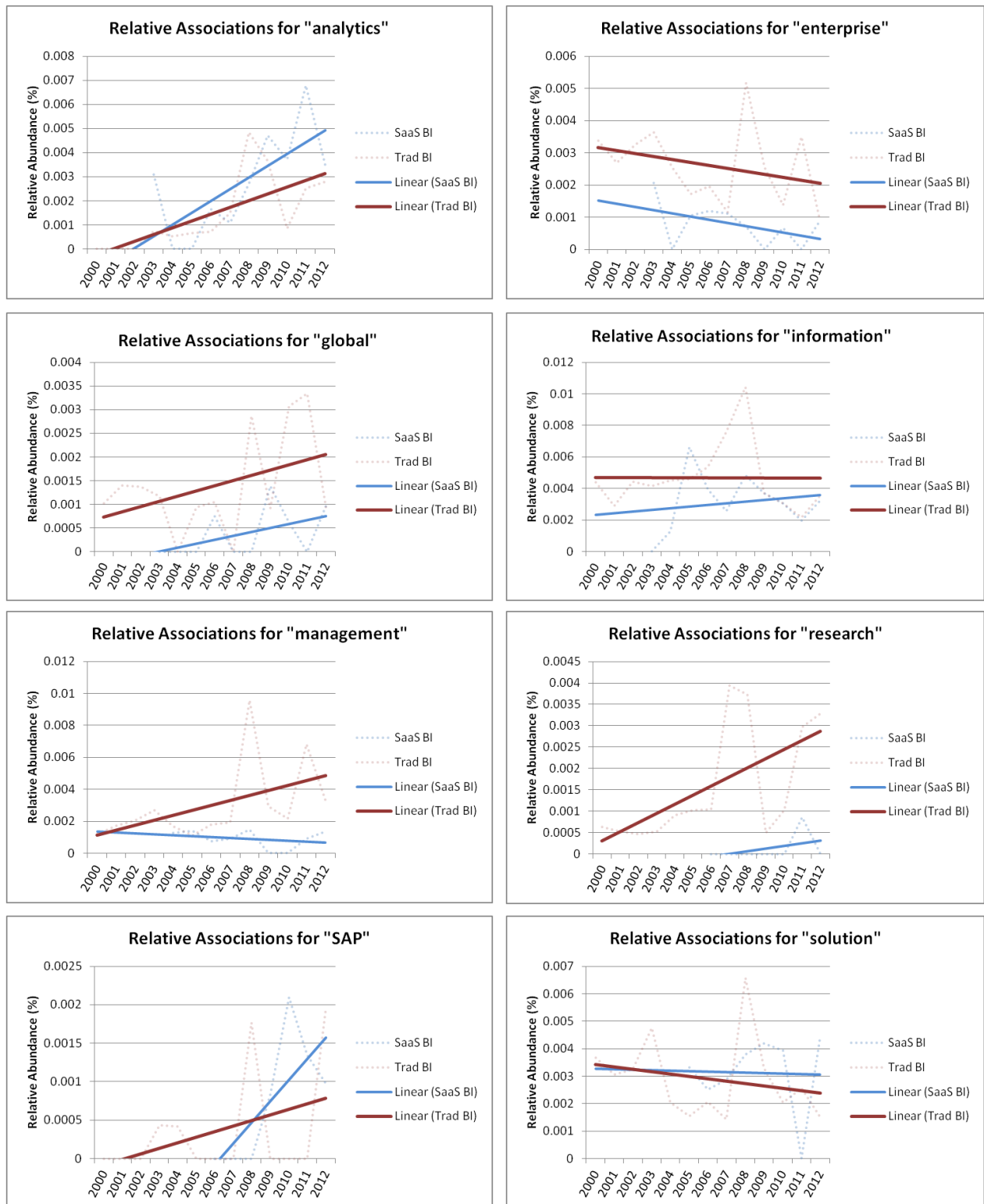
The more prominent association of the words 'partner' and 'SAP' with SaaS BI, can also be explained by the relative novelty of the technology. Partner for example makes reference to the partnership networks that many independent BI vendors in the SaaS BI space are making with hardware (particularly infrastructure) vendors, to provide the platform for their offerings. The association with SAP with SaaS BI can be explained by the fact that SAP acquired Business Objects in 2008, and that SAP has the preeminent position in the total BI market².

Some words were surprisingly absent from the high frequency word lists - 'big data', 'cloud', 'collaborative', 'social', etc. It is possible that these words were too recent in the context of business intelligence, to rank highly enough relative to more established words relating to BI.

² Gartner estimated that the BI analytics and performance management software market surpassed \$(US)12 Billion dollars in 2011. SAP has a market share of 23.6% of the market, while its closest rival Oracle, had 15.6% of the market. <http://www.gartner.com/it/page.jsp?id=1971516>

The trend in relative frequency for each of the 9 highlighted words in figure 4.14 was assessed; the trends for each word's association with SaaS BI and traditional BI are shown graphically in figure 4.15.

Figure 4.15. Relative association of key words with SaaS BI and traditional BI over time.



In the charts in figure 4.15, simple linear trend lines are shown to make it easier to see general trends in relative association. The plots of relative frequency over time for each word's abundance in the SaaS BI and traditional BI corpora is also shown, but have been rendered partially transparent so as not to detract from the overall pattern.

For the time interval studied, there are two broad categories of pattern that can be seen in the charts in figure 4.15, as discussed below.

1. Divergent trends

In this pattern, the association of a given word moves between the SaaS BI and traditional BI groups over time, the plots converging, crossing, and then diverging. This is seen for 'analytics' (χ^2 , 0.02), 'SAP' (χ^2 , 0.06), 'solution' (χ^2 , 0.000009), and 'management' (χ^2 , 0.001). For analytics, SAP, and solution, the shift in association has favored the SaaS BI group, while for management, the converse is true.

For this pattern, the crossover of association may reflect marketing effort on the part of vendors to manipulate the market's perception of their product. The usage of the word 'analytics' in the SaaS BI corpus and the traditional BI corpus is shown as an example. Three text blocks were taken at random from yearly files for SaaS BI and Traditional BI (the years were consistent across the two groups).

For SaaS BI, the text blocks were:

- **2003 - Computer World:** "...three years, we'll see a lot of emphasis away from conventional data warehousing and toward the embedded *analytics*, toward on-demand business intelligence ..."
- **2008 - Computer Business Week:** "...businesses, to provide robust reporting and *analytics* tools that support the sales process. Leveraging Blink Logic's software-as-a-service (SaaS) business intelligence (BI) solution, CRMG provides its customers with a way to take control over their own data, analyze information as required ... "
- **2012 - Wireless News:** "...percent) of respondents' SaaS applications do not currently embed reports, dashboards or *analytics*. As a result, respondents are increasingly turning to SaaS BI to address this gap with 31 percent already implementing this technology, and a further 33 percent planning to in the next two ..."

For Traditional BI, the text blocks were:

- **2000 - PR Newswire Association:** "...parties. The support from such leading organizations as the GartnerGroup, IDC and META Group will be a tremendous asset, as they are one of the leading voices of business intelligence and the evolving *analytics* market."

- **2008 - Investor's Business Daily:** "Business intelligence and *analytics* are a big deal. Anything around decision support is still hot, and that area is very interesting"
- **2012 - CIO Magazine:** "...internal IT capabilities and the products and services your company provides. IT professionals can educate themselves using the fancy business intelligence and *analytics* tools they supply to colleagues in marketing and finance. Javier Polit, CIO of Coca-Cola's Bottling Investments Group, ..."

For the text blocks taken from the SaaS BI corpus, the overall theme is one emphasizing the move to the newer technology of SaaS BI, and linking it with the analytics functions important to business. This is consistent with a newer technology establishing itself in an existing market. For the text blocks taken from the traditional BI corpus, the theme tends to be one of traditional BI already being inextricably linked to the evolving analytics market. This is consistent with traditional BI being in the dominant incumbent position in the market, and also that the technology is still relevant.

It is suspected that words displaying divergent association trends have more diagnostic potential as linguistic markers for being able to distinguish between SaaS BI and traditional BI, as the degree of divergence is increasing over time, and should thus be easier to detect. Cross over events in word associations may prove useful indicators for identifying disruptive technologies, particularly those like that seen for the word "solution", where the association with the incumbent technology is declining while the association with a new technology is climbing. If a critical mass of such word association patterns could be detected, it may represent the tipping point at which a particular technology can be said to have become disruptive. Cross over events may have potential for ex ante identification of disruptive innovations.

2. Approximate parallel trends

In this pattern, the relative association of a word between SaaS BI and traditional BI is relatively stable, although the trend lines can show consistent decay and growth patterns. This overall pattern is seen for 'enterprise' (χ^2 , 0.02), 'global' (χ^2 , 0.04), 'information' (χ^2 , 0.052), and 'research' (χ^2 , 0.053). All of the words in this category show a consistent pattern that favors the association of the word with traditional BI.

It is possible that words in this category displaying growth trends for both SaaS BI and traditional BI ('global', 'research') are seen as equally important by the producers/promoters of the respective technologies, and are being equally used by both in marketing efforts, etc, while words showing decay trends for both groups (such as the word 'enterprise'), are experiencing reduced usage, as their manifest and/or latent meaning is either becoming generic, or is not highly regarded in the marketplace.

The usage of the word 'global' in the SaaS BI corpus and the traditional BI corpus is shown as an example. Three text blocks each were taken at random from yearly files for SaaS BI and Traditional BI (the same years were used for both groups).

For SaaS BI, the text blocks were:

- **2005 - PR Newswire Association:** "The cost to build the kind of infrastructure necessary to deliver our on-demand marketplace for web advertising would be far too prohibitive for a start-up like Etology. We are serving up ads on a *global*, 24x7 basis said Jesse Liu, President of Etology."
- **2008 - Computer Technology Journal:** "...partnership with the Casual Male Retail Group (NASDAQ:CMRG). Oco will support the retailer's *global* expansion by implementing its on-demand BI solution in Casual Male's direct business throughout Europe. Casual Male has been using Oco's BI solution in the ..."
- **2012 - Product News Network:** "Integrating 10 billion hotel rates from more than 400 sources, Rate360 enables hoteliers to best manage property's pricing strategy and ensure rate parity across multiple channels. Program provides up-to-date future rate data along with historical set of rate data on Hotelligence360 cloud-based BI platform. TravelClick, a leading *global* provider of revenue generating solutions for hoteliers, announced today the launch of its Rate360 hotel rate shopping product ..."

For Traditional BI, the text blocks were:

- **2005 - NSTP Computimes:** "... E-business Suite Special Edition to the SME market, with features such as financials, purchasing, inventory management, sales, service, and e-business intelligence. SMEs will have the advantage if they work with *global* players; as such they get more reliable solutions and after-sales support, and better business ..."
- **2008 - Computer Technology Journal:** "Panorama Software, a *global* leader in Proactive Business Intelligence solutions, announced the launch of the Beta version of a new analytics, reporting and data visualization solution for Google applications: ..."
- **2012 - BtOB Media Business:** "...built a financial media company that has *global* reach, possesses deep domain knowledge and derives 50% of its revenue from data and business intelligence advisory services for the mutual fund sector, Casella has built a company that generates about half its revenue from data and business intelligence. Clients such as

Goldman Sachs Group and BlackRock Kelso Capital Corp. split their spending with Asset International between data and marketing, ..."

For the text blocks taken from both the SaaS BI and traditional BI corpora, the usage of the word 'global' is consistent, and relates to its manifest meaning of pertaining to the entire world, the producers and promoters of both technologies both using this word to convey the same meaning.

It is suspected that words displaying association patterns where trends are approximately parallel will have less utility as linguistic markers, but may contribute to an overall pattern which has discriminatory ability.

Finally, for the word 'information', the association trend for traditional BI is remaining stable, but the trend line for association with SaaS BI is trending upwards, and if the trend continues, it is expected that a crossover event will occur in the near future.

5 Discussion

The following sections provide a general discussion of the findings of this study, describe the limitations of the study, and also suggest potential avenues for further research.

5.1 General Discussion

This study looked for linguistic markers for disruptive innovation, using content analysis, and SaaS BI as a test case. No complimentary studies were identified in the academic literature³. Due to the lack of similar studies, this study was undertaken as a preliminary investigation, adopting techniques and ideas from a wide variety of other areas.

Focusing on section 2.1.9.1, which highlighted some of the areas of ongoing debate around Christensen's theory of disruptive innovation, the current study provides some additional insight. Tellis (2006) suggests that a low-end potentially disruptive technology can't be distinguished from an inferior technology that will not go on to be disruptive. In this study, markers indicative of low end disruptions were identified. For example, the vendor attribute set was found to be significantly different between the SaaS BI corpus and the traditional BI Corpus, with higher frequencies being recorded for SaaS BI. When individual vendor numbers were examined over time, it was found that again, there was a statistically significant difference. For traditional BI, the vendor number was

³ For example, the following Google Scholar search strings resulted in 232 results, none directly relevant to the current study: {"content analysis" + "disruptive innovation"}; {"content analysis" + "disruptive innovation" + "linguistic markers"}; {"content analysis" + "disruptive innovation" + "ex ante"}; {"content analysis" + "disruptive innovation" + "SaaS BI"}, {"content analysis" + "disruptive innovation" + "BI SaaS"}.

declining, while for SaaS BI, the vendor number was increasing (and had surpassed that of traditional BI). The Abernathy-Utterback model suggests that an increasing number of small vendors for a new technology, coupled with a decline in existing vendor numbers for an established technology, would be consistent with the market being in the fluid phase; it is during this phase when disruptive technologies typically arise.

Additionally, it was observed that the relative abundance of the industry attribute was significantly different between SaaS BI and traditional BI. The abundance of references to industry attributes had fallen dramatically for traditional BI, whilst rising for SaaS BI. The number of distinct industries followed the same trend, possibly indicating that the market for traditional BI is in decline, while for SaaS BI, it is in a phase of expansion. This observation may be suggestive of the beginnings of a market displacement / disruption.

Govindarajan & Kopalle (2006) provided rigorous statistical analysis of elements of Christensen's theory, and proposed several enhancements to it. One of their contributions was providing evidence that disruptive innovations enter the market from a niche position, and then migrate into the mainstream market. Some evidence supporting this view was found in this study. It was found that the SaaS BI corpus more strongly associated (at a statistically significant level) with the entertainment, mobile, and medical industries, while the traditional BI corpus was more strongly associated with the healthcare, financial, and pharmaceutical industries. On the whole, the entertainment, mobile, and medical industries have largely been underrepresented in traditional BI offerings, and could be considered niche. And it is these newly emergent types of industries that the next wave of business intelligence technologies will target (Chen, Chiang, & Storey, 2012). There was also some evidence supporting the view that the SaaS BI vendors were typically targeting SME's, which adds further credence Govindarajan and Kopalle's idea; it was observed that the words 'enterprise' and 'global' both associated more strongly with the traditional BI corpus, indicating that the vendor discourse for traditional BI was more heavily targeted towards the larger corporate and multinationals, than was the discourse in the SaaS BI corpus.

Yu & Hang (2010) suggest that disruptive innovation does not necessarily mean that incumbents will be displaced by new market entrants. Displacement of an incumbent is dependent upon how the incumbent manages the disruption; some incumbents will falter and be displaced, but others may be able to develop the capability of "ambidexterity", and acquire necessary resources to continue to prosper. Such organizations will not be displaced. Some evidence for this was seen in this study. Only two vendors were represented in both the BI SaaS corpus and the traditional BI corpus; Cognos (IBM) and SAP. Of these two vendors, the relative frequency of SAP was approximately three times that of Cognos.

When looking at just the SaaS BI corpus however, it is seen that of the eight vendors mentioned Sap Business Objects only ranked as the sixth most frequently mentioned vendor. While five of the BI SaaS specific vendors do rank higher than SAP in the BI SaaS corpus, SAP / Business Objects still have the highest installed base in the global BI platform market, and are also in the Gartner leader's quadrant for their Magic Quadrant for Business Intelligence platforms. None of the SaaS BI vendors noted in this study appear in Gartner's Magic Quadrant, presumably because they do not meet the required revenue requirements (\$(US) 15 million in BI related annual software license revenue), and/or all of the required capabilities necessary for inclusion in the report. However, Birst, Oco, and PivotLink were mentioned in the 2011 Magic Quadrant report⁴, while only Birst and PivotLink were mentioned in the 2012 report². In the 2012 Magic Quadrant report, Gartner point out that SAP was the first major traditional BI provider to introduce a SaaS offering. Based on observations to date, it would seem that Yu & Hang's assertion may be correct, as it looks as though SAP is retaining its dominance in the BI market, and is actively competing in the SaaS BI space. In terms of the Dynamic Capabilities Model (Teece, 1996), SAP are displaying 'dynamic capability', which may prevent the newer SaaS BI vendors from displacing their market position. In some ways, SAP's acquisition of Business Objects could be seen as a form of acquiring complementary assets or even as a form of co-operative linkages, which (Rothaermel, 2001) demonstrated could be used by incumbents to protect their dominant position.

5.2 Limitations of Current Study

There were a number of limitations to this study. Firstly, from a methodological perspective, there are limitations with LexisNexis and its ability to robustly provide large full text files (over a standard broadband connection), and its inability to allow extraneous information and duplicate records to be easily culled from the files it outputs, prior to content analysis. This limitation drove the decision to use KWIC output file formats; this may mean that other potentially important words that were not in the same sentence as (or within a 25 word proximity of) the search string used in LexisNexis would be missed. It also precluded the possibility of doing meaningful sentiment analysis.

This limitation could be overcome to some extent, by downloading large full text files from LexisNexis using a high speed broadband connection, such as that provided by KAREN (the Kiwi Advanced Research and Education Network), which Victoria University is a member of, and by pre-

⁴ http://www.board.com/download/press/EN/Gartner_BI_MagicQuadrant_2011.pdf

processing LexisNexis output files with a scripting language such Perl or Python, prior to undertaking content analysis.

Secondly, commercial content analysis tools are prohibitively expensive, and free content analysis tools are understandably limited in the functionality that they offer (and the support available to users of them). This study was restricted by the limitation of the analysis tool chosen, to only be able to output high frequency words to the screen, an inability to process large text files, and its limitation of only showing the top 100 most frequent words.

To a large extent, these limitations could be overcome by using a bespoke approach, with open source tools such as Perl, Python, or R. Hopkins, King, Knowles, & Melendez (2010) describe one such approach utilizing Python and R. There would be an initial steep learning curve which would need to be overcome however. Running analysis tasks on the University's ITS research server, or the Engineering and Computer Sciences computing grid would also add value. It should be pointed out however, that these limitations do not detract from the original intent of the analysis undertaken in this study (which would have remained the same regardless of the tools used), and the results of this investigation demonstrate the potential of the approach adopted by this study.

As a result of the limitations discussed above, a set of choices was made regarding the collection, preparation and processing of data, that may have influenced the findings of this study. For example, one of the choices made was to conduct analysis based upon blocks of text (KWIC blocks) physically centered on the concepts of interest, rather than the full text of articles. As a result of this decision, this study emphasized words that had a textual proximity to the concept of interest. An analysis of the full text articles where the concepts of interest were found may have yielded different results.

Thirdly, a single individual defined the attribute sets, and did the coding of words into the attribute sets. While the coding was simple (and largely based on manifest word meaning), there was still scope for subjectivity and interpretive bias. Future studies should conduct inter-rater reliability assessments. Yet, this study has demonstrated the value of the overall approach used.

From a theoretical basis, one of the limitations of this study was that the discourse of some vendors may be over represented in the corpus. Cukier, Ngwenyama, Bauer, & Middleton (2009) for example, point out that the relative power of the various actors engaged in media discourse cannot be ignored. This study did not explore the source or authorship of the discourse and it is possible that some words may be more strongly associated with certain market participants (vendors, industry analysts, media, users etc.).

5.3 Future Research Opportunities

Given the preliminary nature of this study, there are a number of opportunities for additional research that this study highlighted.

Firstly, there is a large volume of literature regarding whether SaaS technologies in general are considered to be disruptive. While opinion is divided (e.g. Sultan & Bunt-kokhuis (2012) consider SaaS to be disruptive, while Leimeister et al., (2012) do not), there is sufficient published academic opinion to be able to assess SaaS BI as a disruptive innovation along metrics different to those used for content analysis used by this study. Such an analysis may provide useful insight for subsequent content analysis, and for any extensions to this study.

Secondly, due to the infancy of the SaaS BI technology, the volume of practitioner discourse is significantly lower than that which exists for traditional BI. Additional data sources could be investigated to augment data provided by data brokerage houses such as LexisNexis. Social data (twitter, Facebook, LinkedIn), may provide additional insights, as might Blogs, Micro-blogs, discussion forums, etc. These data sources all provide data amenable to investigation via content analysis. Tayal & Komaragiri (2009) for example, used content analysis of micro-blogs to predict market performance of Google. Incorporating industry analyst reports would also provide additional insight.

Thirdly, sentiment analysis was not undertaken as part of this study, but may have proven useful in determining perceptions regarding whether traditional BI was meeting current consumer needs, and additionally, what the market perception of SaaS BI was. Tracking sentiment may contribute to the predictive capacity of content analysis in identifying disruptive innovations ex ante. Social media may prove most useful for this task (Bollen, Mao, & Zeng (2011), Asur & Huberman (2010)). Undertaking analysis which compared the various data sources may also be insightful. For example, can some sources be used to predict changes in sentiment among other sources? Are some authors more influential than others? Is sentiment consistent across all sources, or does it cluster? All of these questions could provide important insights into how discourse parallels or anticipates disruptive technological innovations.

6 Conclusion

The scope of this project was to investigate whether useful linguistic characteristics could be identified in ICT practitioner discourse, that could be used to discriminate between SaaS BI (a candidate disruptive innovation) and traditional BI (the incumbent technology). Content analysis was the technique chosen to undertake this analysis, and analysis was performed using attribute

sets (capturing both manifest and some latent meaning of component words), and also individual relative word associations with both SaaS BI and traditional BI corpora. The analysis based upon attribute set usage provided evidence that attribute sets are stable enough to be used for this purpose, and was suggestive of the fact that SaaS BI could be a disruptive technology, while analysis of individual word associations with both SaaS BI and traditional BI corpora provided a number of word association patterns that could discriminate between SaaS BI and traditional BI. These patterns (especially the crossover events, where the pattern of word associations switch between incumbent and new technology) may be transferable to other technologies, and may be able to provide an indication that a technology innovation is, or is about to become, disruptive. It is possible that such cross over events may have predictive capability.

The research question addressed by this study was whether discriminative markers can be identified in practitioner discourse that could be used to predict whether a technological innovation was going to be disruptive. The results of this investigation suggest the answer is yes.

7 References

- Abernathy, W., & Utterback, J. (1978). Patterns of Industrial Innovation. *Technology Review (MIT)*, 40–47.
- Abrahamson, E., & Hambrick, D. C. (1997). Attentional Homogeneity in Industries: The Effect of Discretion. *Journal of Organizational Behavior*, 18(S1), 513–532.
- Adner, R. (2002). When are Technologies Disruptive? A Demand-Based View of the Emergence of Competition. *Strategic Management Journal*, 23(8), 667–688. doi:10.1002/smj.246
- Afuah, A. (2009). *Strategic Innovation. New Game Strategies for Competitive Advantage* (1st ed.). New York: Routledge.
- Afuah, A., & Utterback, J. (1997). Responding to Structural Industry Changes: A Technological Evolution Perspective. *Industrial & Corporate Change*, 6(1), 183–202.
- Andreu, L., Caraganis, H., & Okazaki, S. (2011). Social Networking Sites for Higher Education: A Content Analysis Among US Top MBA Programs. *4th International Conference of Education, Research and Innovations (ICERI2011 Proceedings)* (pp. 18–21). Madrid: ICERI.
- Armbrust, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R., Konwinski, A., Lee, G., et al. (2010). Above the Clouds: A Berkeley View of Cloud Computing. *Communications of the ACM*, 53(4), 50–58.
- Assink, M. (2006). Inhibitors of Disruptive Innovation Capability: a Conceptual Model. *European Journal of Innovation Management*, 9(2), 215–233. doi:10.1108/14601060610663587
- Asur, S., & Huberman, B. A. (2010). Predicting the Future with Social Media. *arXiv*. Ithaca.
- Barney, J. B. (1997). On Flipping Coins and Making Technology Choices: Luck as an Explanation of Technological Foresight and Oversight. In R. Garud, P. Nayyar, & Z. Shapira (Eds.), *Technological Innovation: Oversight and Foresights* (pp. 13–19). New York: Cambridge University Press.
- Barras, R. (1986). Towards a Theory of Innovation in Services. *Research Policy*, 15(4), 161–173.
- Barras, R. (1990). Interactive Innovation in Financial and Business Services: The Vanguard of the Service Revolution. *Research Policy*, 19(3), 215–237.
- Bejan, A., & Lorente, S. (2011). The Constructal Law Origin of the Logistics S Curve. *Journal of Applied Physics*, 110(2). doi:10.1063/1.3606555
- Bejan, Adrian, & Lorente, S. (2010). The Constructal Law of Design and Evolution in Nature. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 365(1545), 1335–47. doi:10.1098/rstb.2009.0302

- Bejan, Adrian, & Lorente, S. (2012). The physics of spreading ideas. *International Journal of Heat and Mass Transfer*, 55(4), 802–807. doi:10.1016/j.ijheatmasstransfer.2011.10.029
- Bengston, D. N., Potts, R. S., Fan, D. P., & Goetz, E. G. (2005). An Analysis of the Public Discourse About Urban Sprawl in the United States: Monitoring Concern About a Major Threat to Forests. *Forest Policy and Economics*, 7(5), 745–756. doi:10.1016/j.forpol.2005.03.010
- Bennett, K., Layzell, P., Budgen, D., Brereton, P., Macaulay, L., & Munro, M. (2000). Service-based Software: The Future for Flexible Software. *Seventh Asia-Pacific Software Engineering Conference* (pp. 214–221). doi:10.1109/APSEC.2000.896702
- Bezemer, C., & Zaidman, A. (2010). Multi-Tenant SaaS Applications: Maintenance Dream or Nightmare? *Fourth International Joint ERCIM/IWPSE Symposium on Software Evolution* (pp. 88–92). ACM.
- Bitterer, A., Schlegel, K., Hostmann, B., Gassman, B., Rayner, N., Chandler, N., Beyer, A., et al. (2007). *Hype Cycle for Business Intelligence and Performance Management, 2007*. Stamford, CT.
- Bollen, J., Mao, H., & Zeng, X. (2011). Twitter Mood Predicts the Stock Market. *Journal of Computational Science*, 2(1), 1–8. doi:10.1016/j.jocs.2010.12.007
- Bower, J., & Christensen, C. (1995). Disruptive Technologies: Catching the Wave. *Harvard Business Review*, 73(1), 43–53.
- Bowling, S. R., Khasawneh, M. T., Kaewkuekool, S., & Cho, B. R. (2009). A Logistic Approximation to the Cumulative Normal Distribution. *Journal of Industrial Engineering and Management*, 2(1), 114–127. doi:10.3926/jiem.2009.v2n1.p114-127
- Brown, R. (1992). Managing the “S” Curves of Innovation. *The Journal of Consumer Marketing*, 9(1), 61–72.
- Bunnell, T. G., & Coe, N. M. (2001). Spaces and scales of innovation. *Progress in Human Geography*, 25(4), 569–589. doi:10.1191/030913201682688940
- Carley, K. (1993). Coding Choice for Textual Analysis: A Comparison of Content Analysis and Map Analysis. *Sociological Methodology*, 23, 75–126.
- Carley, K. (1997). Extracting Team Mental Models Through Textual Analysis. *Journal of Organizational Behavior*, 18(S1), 533–558. doi:10.1002/(SICI)1099-1379(199711)18:1+<533::AID-JOB906>3.3.CO;2-V
- Chandy, R. K., & Tellis, G. J. (2000). The Incumbent’s Curse? Incumbency, Size, and Radical Product Innovation. *Journal of Marketing*, 64(3), 1–17.
- Chang, S.-B., Lai, K., & Chang, S.-M. (2009). Exploring Technology Diffusion and Classification of Business Methods: Using the Patent Citation Network. *Technological Forecasting & Social Change*, 76, 107–117. doi:doi:10.1016/j.techfore.2008.03.014
- Chee, T., Chan, L.-K., Chuah, M.-H., Tan, C.-S., Wong, S.-F., Yeoh, W., & Rahman, A. (2009). Business Intelligence Systems: State of the Art Review and Contemporary

Applications. *Symposium on Progress in Information & Communication Technology* (pp. 96–101).

- Chen, H., Chiang, R., & Storey, V. C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. *MIS Quarterly*, 36(4), 1–24.
- Chesbrough, H., & Crowther, A. K. (2006). Beyond High Tech: Early Adopters of Open Innovation in Other Industries. *R&D Management*, 36(3), 229–236. doi:10.1111/j.1467-9310.2006.00428.x
- Christensen, C. (1992a). Exploring the Limits of the Technology S-curve. Part 1: Component Technologies. *Production and Operations Management*, 1(4), 334–357. doi:10.1111/j.1937-5956.1992.tb00001.x
- Christensen, C. (1992b). Exploring the Limits of the Technology S-curve. Part 2: Architectural Technologies. *Production and Operations Management*, 1(4), 358–366.
- Christensen, C. (1997). *The Innovator's Dilemma. When New Technologies Cause Great Firms to Fail*. Boston, MA.: Harvard Business School Press.
- Christensen, C. (2006). The Ongoing Process of Building a Theory of Disruption. *Journal of Product Innovation Management*, 26(1), 39–55.
- Christensen, C., & Bower, J. L. (1996). Customer Power, Strategic Investment, and the Failure of Leading Firms. *Strategic Management Journal*, 17(3), 197–218.
- Christensen, C., & Raynor, M. (2003). *The Innovator's Solution: Creating and Sustaining Successful Growth*. Boston, MA: Harvard Business School Press.
- Christensen, C., & Rosenbloom, R. S. (1995). Explaining the Attacker's Advantage: Technological Paradigms, Organizational Dynamics, and the Value Network. *Research Policy*, 24(2), 233–257.
- Claude-gaudillat, V., & Quélin, B. V. (2006). Innovation, New Market and Governance Choices of Entry: The Internet Brokerage Market Case. *Industry & Innovation*, 13(2), 173–187. doi:10.1080/13662710600684324
- Cole, T. J., & Green, P. J. (1992). Smoothing Reference Centile Curves: The LMS Method and Penalized Likelihood. *Statistics in medicine*, 11(10), 1305–19.
- Cukier, W., Ngwenyama, O., Bauer, R., & Middleton, C. (2009). A Critical Analysis of Media Discourse on Information Technology: Preliminary Results of a Proposed Method for Critical Discourse analysis. *Information Systems Journal*, 19(2), 175–196. doi:10.1111/j.1365-2575.2008.00296.x
- Cukier, W., Trenholm, S., Carl, D., & Gekas, G. (2011). Social Entrepreneurship: A Content Analysis. *Journal of Strategic Innovation and Sustainability*, 7(1), 99–120.
- Damanpour, F., & Gopalakrishnan, S. (2001). The Dynamics of the Adoption of Product and Process Innovations in Organizations. *Journal of Management Studies*, 38(1), 45–65.
- Danneels, E. (2004). Disruptive Technology Reconsidered: A Critique and Research Agenda. *Journal of Product Innovation Management*, (21), 246–258.

- Davenport, T. H., & Harris, J. G. (2007). *Competing on Analytics. The New Science of Winning*. Boston, MA: Harvard Business School Press.
- Davis, J., Miller, G. J., & Russell, A. (2006). *Information Revolution. Using the Information Evolution Model to Grow your Business*. New Jersey: John Wiley & Sons, Inc.
- DeTienne, D. R., & Koberg, C. S. (2002). The Impact of Environmental and Organizational Factors on Discontinuous Innovation Within High-Technology Industries. *IEEE Transactions on Engineering Management*, 49(4), 352–364. doi:10.1109/TEM.2002.806719
- Dewar, R. D., & Dutton, J. (1986). The Adoption of Radical and Incremental Innovations: An Empirical Analysis. *Management Science*, 32(11), 1422–1433. doi:10.1287/mnsc.32.11.1422
- Dosi, G. (1982). Technological Paradigms and Technological Trajectories. *Research Policy*, 11(3), 147–162.
- Dosi, G. (1988). Sources, Procedures, and Microeconomic Effects of Innovation. *Journal of Economic Literature*, 26(3), 1120–1171.
- Druehl, C. T., & Schmidt, G. M. (2008). A Strategy for Opening a New Market and Encroaching on the Lower End of the Existing Market. *Production and Operations Management*, 17(1), 44–60. doi:10.3401/poms.1070.0002
- Duriau, V. J., Reger, R. K., & Pfarrer, M. D. (2007). A Content Analysis of the Content Analysis Literature in Organization Studies: Research Themes, Data Sources, and Methodological Refinements. *Organizational Research Methods*, 10(1), 5–34.
- Elbashir, M. Z., Collier, P. a., & Sutton, S. G. (2011). The Role of Organizational Absorptive Capacity in Strategic Use of Business Intelligence to Support Integrated Management Control Systems. *The Accounting Review*, 86(1), 155–184. doi:10.2308/accr.00000010
- Elo, S., & Kyngäs, H. (2008). The Qualitative Content Analysis Process. *Journal of Advanced Nursing*, 62(1), 107–15. doi:10.1111/j.1365-2648.2007.04569.x
- Fiss, P. C., & Hirsch, P. M. (2005). The Discourse of Globalization: Framing and Sensemaking of an Emerging Concept. *American Sociological Review*, 70(1), 29–52. doi:10.1177/000312240507000103
- Folkestad, J., & Gonzalez, R. (2010). Teamwork for Innovation: A Content Analysis of the Highly Read and Highly Cited Literature on Innovation. *Advances in Developing Human Resources*, 12(1), 115–136. doi:10.1177/1523422310365486
- Foster, R. N. (1985). Timing Technological Transitions. *Technology In Society*, 7(2-3), 127–141.
- Garcia, R., & Calantone, R. (2002). A critical look at technological innovation typology and innovativeness terminology: a literature review. *Journal of Product Innovation Management*, 19(2), 110–132. doi:10.1111/1540-5885.1920110
- Gartner. (2011). *Hype Cycle for Business Intelligence , 2011 (G00216086)*. Stamford, CT.

- Gatlin, J. (1999). *Bill Gates: The Path to the Future*. New York: Avon Books.
- Gephart, R. (1993). The Textual Approach: Risk and Blame in Disaster Sensemaking. *Academy of Management Journal*, 36(6), 1465–1514. doi:10.2307/256819
- Gephart, R. (1997). Hazardous Measures: an Interpretive Textual Analysis of Quantitative Sensemaking During Crises. *Journal of Organizational Behavior*, 18(S1), 583–622. doi:10.1002/(SICI)1099-1379(199711)18:1+<583::AID-JOB908>3.0.CO;2-T
- Ghosh, A. (2012). Is there an S-curve Relationship Between U.S. Trade Balance and Terms of Trade? An Analysis Across Industries and Countries. *Economics Bulletin*, 32(1), 325–337.
- Gloia, D., & Chittipeddi, K. (1991). Sensemaking and Sensegiving in Strategic Change Initiation. *Strategic Management Journal*, 12(6), 433–448.
- Gonzales, M. L., Udo, G., Bagchi, K., & Kirs, P. (2011). Diffusion of Business Intelligence and Data Warehousing □: An Exploratory Investigation of Research and Practice University of Texas at El Paso University of Texas at El Paso. *Proceedings of the 44th Hawaii International Conference on System Sciences* (pp. 1–9). ICCSN IEEE.
- Gort, M., & Klepper, S. (1982). Time Paths in the Diffusion of Product Innovations. *The Economic Journal*, 92(367), 630–653.
- Govindarajan, V., & Kopalle, P. (2004). How Legacy Firms Can Introduce Radical and Disruptive Innovations: Theoretical and Empirical Analyses. *Academy of Management Proceedings* (p. A1). New York: Academy of Management.
- Govindarajan, V., & Kopalle, P. K. (2006). The Usefulness of Measuring Disruptiveness of Innovations Ex Post in Making Ex Ante Predictions. *Journal of Product Innovation Management*, 23, 12–18.
- Grimmer, J., & Stewart, B. M. (2012). Text as Data □: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. Stanford, CA. Retrieved from <http://www.stanford.edu/~jgrimmer/>
- Haramé, D. L., & Meyerson, B. S. (2001). The Early History of IBM 's SiGe Mixed Signal Technology. *IEEE Transactions on Electronic Devices*, 48(11), 2555–2567.
- Henderson, R., & Clark, K. (1990). Architectural Innovation: The Reconfiguration of Existing Product Technologies and the Failure of Established Firms. *Administrative Science Quarterly*, 35(1), 9–30.
- Hienerth, C., & Lettl, C. (2011). Innovations to Become Standard Equipment in the Industry □: Community Pull Effects. *Journal of Product Innovation Management*, 28(S1), 175–195.
- Hopkins, D. J., & King, G. (2010). A Method of Automated Nonparametric Content Analysis for Social Science. *American Journal of Political Science*, 54(1), 229–247. doi:10.1111/j.1540-5907.2009.00428.x
- Hopkins, D., King, G., Knowles, M., & Melendez, S. (2010). *ReadMe* □: *Software for Automated Content Analysis*. Institute for Quantitative Social Science.

- Howson, C. (2008). *Successful Business Intelligence: Secrets to Making BI a Killer App*. New York: McGraw-Hill.
- Hsieh, H.-F., & Shannon, S. E. (2005). Three Approaches to Qualitative Content Analysis. *Qualitative Health Research*, 15(9), 1277–88. doi:10.1177/1049732305276687
- Huettenmueller, R. (2006). *Business Calculus Demystified*. New York: McGraw-Hill.
- Hysong, S. J., Sawhney, M. K., Wilson, L., Sittig, D. F., Esquivel, A., Watford, M., Davis, T., et al. (2009). Improving Outpatient Safety Through Effective Electronic Communication: A Study Protocol. *Implementation Science*, 4, 62. doi:10.1186/1748-5908-4-62
- Inauen, M., & Schenker-Wicki, A. (2012). Fostering Radical Innovations With Open Innovation. *European Journal of Innovation Management*, 15(2), 212–231. doi:10.1108/14601061211220977
- Jiang, Y., Shang, J., & Liu, Y. (2010). Maximizing customer satisfaction through an online recommendation system: A novel associative classification model. *Decision Support Systems*, 48(3), 470–479. doi:10.1016/j.dss.2009.06.006
- Kabanoff, B. (1997). Computers Can Read as Well as Count: Computer-aided Text Analysis in Organizational Research. *Journal of Organizational Behavior*, 18(S1), 507–511.
- Kabanoff, B., & Holt, J. (1996). Changes in the Espoused Values of Australian Organizations 1986 -1990. *Journal of Organizational Behavior*, 17(1), 201–219.
- Kirilenko, A. P., & Stepchenkova, S. (2012). Climate Change Discourse in Mass Media: Application of Computer-Assisted Content Analysis. *Journal Of Environmental Studies And Sciences*, 2(4), 178–191. doi:10.1007/s13412-012-0074-z
- Klepper, S., & Graddy, E. (1990). The Evolution of New Industries and the Determinants of Market Structure. *The RAND Journal of Economics*, 21(1), 27–44. doi:10.2307/2555491
- Kongparakul, S., Ng, F. T. T., & Rempel, G. L. (2011). Metathesis Hydrogenation of Natural Rubber Latex. *Applied Catalysis A: General*, 405(1-2), 129–136. doi:10.1016/j.apcata.2011.07.039
- Kostoff, R. N., Boylan, R., & Simons, G. R. (2004). Disruptive Technology Roadmaps. *Technological Forecasting and Social Change*, 71(1-2), 141–159. doi:10.1016/S0040-1625(03)00048-9
- Krippendorff, K. (2012). *Content Analysis: An Introduction to its Methodology* (3rd Editio.). Thousand Oaks CA: Sage Publications.
- Kumar, P. (2012). Impact of Business Intelligence Systems in Indian Telecom Industry. *Business Intelligence Journal*, 5(2), 358–366.
- Lavalle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2011). Big Data , Analytics and the Path from Insights to Value. *MIT Sloan Management Review*, 52(2), 20–31.

- Lee, R. P., & Chen, Q. (2009). The Immediate Impact of New Product Introductions on Stock Price: The Role of Firm Resources and Size. *Journal of Product Innovation Management*, 26(1), 97–107. doi:10.1111/j.1540-5885.2009.00337.x
- Lee, Z., Gosain, S., & Im, I. (1999). Topics of Interest in IS: Evolution of Themes and Differences Between Research and Practice. *Information & Management*, 36(5), 233–246. doi:10.1016/S0378-7206(99)00022-1
- Leimeister, Stefanie; Riedl, Christoph; Böhm, Markus; Krcmar, H. (2012). The Business Perspective of Cloud Computing□: Actors , Roles , and Value Networks. *Proceedings of 18th European Conference on Information Systems* (p. Paper 56).
- Levine, D. S. (2012). Neural Dynamics of Affect, Gist, Probability, and Choice. *Cognitive Systems Research*, 15-16, 57–72. doi:10.1016/j.cogsys.2011.07.002
- Lissack, M. (1998). Concept Sampling: A New Twist for Content Analysis. *Organizational Research Methods*, 1(4), 484–504.
- Liyang, T., Zhiwei, N., Zhangjun, W., & Li, W. (2011). A Conceptual Framework for Business Intelligence as a Service (SaaS BI). *Fourth International Conference on Intelligent Computation Technology and Automation* (pp. 1025–1028). IEEE Comput. Soc. doi:10.1109/ICICTA.2011.541
- Lynn, G. S., Morone, J. G., & Paulson, A. S. (1996). Marketing and Discontinuous Innovation: The Probe and Learn Process. *California Management Review*, 38(3), 8–37.
- MacDonald, L. D., & Richardson, A. J. (2011). Does Academic Management Accounting Lag Practice? A Cliometric Study. *Accounting History*, 16(4), 365–388. doi:10.1177/1032373211417989
- Mantzoukas, S. (2009). The Research Evidence Published in High Impact Nursing Journals Between 2000 and 2006: A Quantitative Content Analysis. *International Journal of Nursing Studies*, 46(4), 479–89. doi:10.1016/j.ijnurstu.2008.12.016
- Martens, B., Poeppelbuss, J., & Teuteberg, F. (2011). Understanding the Cloud Computing Ecosystem□: Results from a Quantitative Content Analysis. *Proceedings of the 10th International Conference on Wirtschaftsinformatik*. Zürich.
- Martin, R. G., & Lubow, S. H. (2011). The Gravo-Magneto Limit Cycle in Accretion Disks. *The Astrophysical Journal Letters*, 740(1), L6. doi:10.1088/2041-8205/740/1/L6
- Mascitelli, R. (2000). From Experience: Harnessing Tacit Knowledge to Achieve Breakthrough Innovation. *Journal of Product Innovation Management*, 17(3), 179–193.
- Mayring, P. (2000). Forum□: Qualitative Social Research (Sozialforschung). *Forum, Qualitative Social Research*, 1(2), 20.
- Messinger, A. (2012). Teaching Content Analysis through Harry Potter. *Teaching Sociology*, 40(2), 1–8. doi:10.1177/0092055X12445461
- Mitchell, W. (1989). Whether and When? Probability and Timing of Incumbents' Entry into Emerging Industrial Subfields. *Administrative Science Quarterly*, 34(2), 208–230.

- Moretti, F., van Vliet, L., Bensing, J., Deledda, G., Mazzi, M., Rimondini, M., Zimmermann, C., et al. (2011). A Standardized Approach to Qualitative Content Analysis of Focus Group Discussions From Different Countries. *Patient Education and Counseling*, 82(3), 420–8. doi:10.1016/j.pec.2011.01.005
- Mossholder, K., Settoon, R., Harris, S., & Armenakis, A. (1995). Measuring Emotion in Open Ended Survey Responses: An Application of Textual Data Analysis. *Journal of Management*, 21(2), 335–355.
- Mueller, D., & Tilton, J. (1969). Research and Developments Costs as a Barrier to Entry. *Canadian Journal of Economics*, 2(4), 570–579.
- Myers, D., Sumpter, C., & Walsh, S. (2002). Guest Editorial. A Practitioner's View: Evolutionary Stages of Disruptive Technologies. *IEEE Transactions on Engineering Management*, 49(4), 322–329.
- National Research Council. (2010). *Persistent Forecasting of Disruptive Technologies*. Washington, D.C.: National Academies Press.
- Nelson, R., & Winter, S. (1977). In Search of Useful Theory of Innovation. *Research Policy*, 6(1), 36–76.
- Neuendorf, K. (2011). Content Analysis—A Methodological Primer for Gender Research. *Sex Roles*, 64(3-4), 276–289. doi:10.1007/s11199-010-9893-0
- Ouf, S., Nasr, M., Amr, M., Mosaad, M., Kamal, K., Mostafa, F., Said, R., et al. (2011). Business Intelligence Software as a Service (SAAS). *3rd International Conference on Communication Software and Networks* (pp. 641–649). ICCSN IEEE. doi:10.1109/ICCSN.2011.6014350
- Owlia, M. S. (2010). A Framework for Quality Dimensions of Knowledge Management Systems. *Total Quality Management & Business Excellence*, 21(11), 1215–1228. doi:10.1080/14783363.2010.529351
- Paap, J., & Katz, R. (2004). Anticipating Disruptive Innovation. *Research Technology Management*, 47(5), 13–22.
- Pollock, N. (2009). Constructing Technology Markets: How Industry Analysts Organise Technological Fields. *Proceedings of ALPIS: itAIS, Italy. Sprouts Working Papers on Information Systems*, 9(10) (Vol. 9).
- Pollock, N., & Williams, R. (2011). Who Decides the Shape of Product Markets? The Knowledge Institutions that Name and Categorise New Technologies. *Information and Organization*, 21(4), 194–217. doi:10.1016/j.infoandorg.2011.08.001
- Punch, K. (2005). *Introduction to Social Research: Quantitative and Qualitative Approaches* (2nd ed.). Thousand Oaks, CA: Sage Publications Ltd.
- Pyka, A., & Andersen, E. S. (2012). Introduction: Long Term Economic Development – Demand, Finance, Organization, Policy and Innovation in a Schumpeterian Perspective. *Journal of Evolutionary Economics*, 22(4), 621–625. doi:10.1007/s00191-012-0279-z

- Ransbotham, S., & Mitra, S. (2011). The Impact of Immediate Disclosure on Attack Diffusion and Volume. *Workshop on the Economics of Information Security*. Fairfax, VA.
- Rich, T. (2012). Deciphering North Korea's Nuclear Rhetoric: An Automated Content Analysis of KCNA News. *Asian Affairs: An American Review*, 39(2), 73–89. doi:10.1080/00927678.2012.678128
- Rivard, S., Hautes, I., & Raymond, L. (1997). Development of a Measure to Assess the Quality of User- Developed Applications. *The DATA BASE for Advances in Information Systems*, 28(3), 44–58.
- Rivas, R. M. (2007). A Study Of Corporate Entrepreneurship And Firm Performance In The Computer Industry During Two Technological Disruptions. *Review of Business Information Systems*, 11(3), 65–74.
- Roberts, C. W. (1989). Other Than Counting Words: A Linguistic Approach to Content Analysis. *Social Forces*, 68(1), 147. doi:10.2307/2579224
- Rogers, E. (2003). *Diffusion of Innovations* (5th ed.). New York: Free Press.
- Ronda-Pupo, G. A., & Angel, L. (2012). Dynamics of the Evolution of the Strategy Concept 1962 – 2008 □: A Co-Word Analysis. *Strategic Management Journal*, 33(2), 162–188. doi:10.1002/smj
- Rothaermel, F. T. (2001). Incumbent's Advantage Through Exploiting Complementary Assets Via Interfirm Cooperation. *Strategic Management Journal*, 22(6-7), 687–699. doi:10.1002/smj.180
- Rothaermel, F. T. (2002). Technological Discontinuities and Interfirm Cooperation □: What Determines a Startup ' s Attractiveness as Alliance Partner □? *IEEE Transactions on Engineering Management*, 49(4), 388–397.
- Schmidt, G. M. (2004). Low-End and High-End Encroachment Strategies for New Products. *International Journal of Innovation Management*, 8(2), 167–191.
- Schmidt, G. M., & Druehl, C. T. (2008). When Is a Disruptive Innovation Disruptive? *Journal of Product Innovation Management*, 25(4), 347–369.
- Shane, S. (2008). *Technology Strategy for Managers and Entrepreneurs* (1st ed.). New York: Prentice Hall.
- Sheikh, R. A. (2011). SaaS BI □: Sustainable Business Intelligence Solution for SMB ' S. *International Journal of Research in Finance & Marketing*, 1(3), 1–11.
- Sima, C. S., Panageas, K. S., Heller, G., & Schrag, D. (2010). Analytical Strategies for Characterizing Chemotherapy Diffusion with Patient-Level Population-Based Data. *Applied Health Economics and Health Policy*, 8(1), 37–51. doi:10.2165/11314090-000000000-00000
- Simons, S. S., & Chow, C. C. (2012). The Road Less Traveled: New Views of Steroid Receptor Action from the Path of Dose-Response Curves. *Molecular and Cellular Endocrinology*, 348(2), 373–82. doi:10.1016/j.mce.2011.05.030

- Sjovaag, H., & Stavelin, E. (2012). Web Media and the Quantitative Content Analysis: Methodological Challenges in Measuring Online News Content. *Convergence: The International Journal of Research into New Media Technologies*, 18(2), 215–229. doi:10.1177/1354856511429641
- Slapin, J. B., & Proksch, S.-O. (2008). A Scaling Model for Estimating Time-Series Party Positions from Texts. *American Journal of Political Science*, 52(3), 705–722. doi:10.1111/j.1540-5907.2008.00338.x
- Soojung, K., & Snaghee, O. (2009). Users' Relevance Criteria for Evaluating Answers in a Social Q & A Site. *Journal of the American Society for Information Science and Technology*, 60(4), 716–727. doi:10.1002/asi
- Spiegel, M. (1992). *Schaum's Outline of Theory and Problems of Statistics* (2nd ed.). London: McGraw-Hill Book Company.
- Stemler, S. (2001). An Overview of Content Analysis. *Practical Assessment, Research & Evaluation*, 7(17), 1–9.
- Stipic, A., & Bronzin, T. (2012). How Cloud Computing is (not) Changing the Way We Do BI. *MIPRO Proceedings of the 35th International Convention* (pp. 1574–1582). Opatija, Croatia: IEEE Proceedings.
- Stoneman, P., & Diederer, P. (1994). Technology Diffusion and Public Policy. *The Economic Journal*, 104(425), 918–930.
- Suarez, F. F., & Utterback, J. (1995). Dominant Designs and the Survival of Firms. *Strategic Management Journal*, 16, 415–430.
- Sultan, N., & Bunt-kokhuis, S. V. D. (2012). Organisational Culture and Cloud Computing: Coping with a Disruptive Innovation. *Technology Analysis & Strategic Management*, 24(2), 167–179.
- Swanson, E. B. (2010). Consultancies and Capabilities in Innovating with IT. *The Journal of Strategic Information Systems*, 19(1), 17–27. doi:10.1016/j.jsis.2009.12.001
- Tangpong, C. (2011). Content Analytic Approach to Measuring Constructs in Operations and Supply Chain Management. *Journal of Operations Management*, 29(6), 627–638. doi:10.1016/j.jom.2010.08.001
- Tayal, D., & Komaragiri, S. (2009). Comparative Analysis of the Impact of Blogging and Micro-blogging on Market Performance. *International Journal on Computer Science and Engineering*, 1(3), 176–182.
- Taylor, C. A., Francony, J. J., Beiting, S. R., Ritter, J., & Clutter, J. E. (2011). A Qualitative Content Analysis of Local School Wellness Policies for Ohio Schools. *Infant, Child, & Adolescent Nutrition*, 3(1), 9–15. doi:10.1177/1941406410396080.
- Teece, D. (1986). Profiting from Technological Innovation: Implications for Integration, Collaboration, Licensing and Public Policy. *Research Policy*, 15(6), 285–305.
- Teece, D. (1996). Firm Organization, Industrial Structure, and Technological Innovation. *Journal of Economic Behavior & Organization*, 31(2), 193–224.

- Teece, D., & Pisano, G. (1994). *The Dynamic Capabilities of Firms: An Introduction*. Laxenburg (Austria).
- Teece, D., Pisano, G., & Shuen, A. (1997). Dynamic Capabilities and Strategic Management. *Strategic Management Journal*, 18(7), 509–533.
- Tellis, G. J. (2006). Disruptive Technology or Visionary Leadership? *Journal of Product Innovation Management*, 23(1), 34–38. doi:10.1111/j.1540-5885.2005.00179.x
- Thomond, P., Herzberg, T., & Lettice, F. (2003). Disruptive Innovation: Removing the Innovators' Dilemma. *Knowledge into Practice - British Academy of Management Annual Conference*. Harrogate, UK.
- Tripsas, M. (2007). Customer Preference Discontinuities: A Trigger for Radical Technological Change. *Managerial and Decision Economics*, 29, 79–97. doi:10.1002/mde
- Tushman, M. L., & Anderson, P. (1986). Technological Discontinuities and Organizational Environments. *Administrative Science Quarterly*, 31(3), 439–465.
- Usrey, M., & Dooley, K. (1996). The Dimensions of Software Quality. *Quality Management Journal*, 3(3), 67–86.
- Usrey, M., & Dooley, K. (1998). The Measurement of Consumer Attitudes Concerning Software Quality. *Quality Management Journal*, 5(2), 42–57.
- Utterback, J. (1994). *Mastering the Dynamics of Innovation*. Cambridge, MA: Harvard Business School Press.
- Utterback, J. M., & Abernathy, W. J. (1975). A Dynamic Model of Process and Product Innovation. *OMEGA: The International Journal of Management Science*, 3(6), 639–656.
- Van Orden, J., van der Rhee, B., & Schmidt, G. M. (2011). Encroachment Patterns of the “Best Products” from the Last Decade. *Journal of Product Innovation Management*, 28(5), 726–743. doi:10.1111/j.1540-5885.2011.00834.x
- Veryzer, R. W. (1998). Discontinuous Innovation and the New Product Development Process. *Journal of Product Innovation Management*, 15(4), 304–321.
- Vo, A. T. (2012). Visualizing Context Through Theory Deconstruction: A Content Analysis of Three Bodies of Evaluation Theory Literature. *Evaluation and Program Planning*, 35(In Press). doi:10.1016/j.evalprogplan.2012.03.013
- Webb, T. J., Bengston, D. N., & Fan, D. P. (2008). Forest Value Orientations in Australia: An Application of Computer Content Analysis. *Environmental Management*, 41(1), 52–63. doi:10.1007/s00267-007-9011-4
- Weber, R. (1990). *Basic Content Analysis* (2nd ed.). Thousand Oaks, CA: Sage Publications.
- William, L. (2006). Innovation Rules! *Research Technology Management*, 49(2), 8 – 14.
- Winsor, C. P. (1932). The Gompertz Curve as a Growth Curve. *Proceedings of the National Academy of Sciences*, 18(1), 1 – 8.

- Wolfe, R., Gephart, R., & Johnson, T. (1993). Computer-Facilitated Qualitative Data Analysis: Potential Contributions to Management Research. *Journal of Management*, 19(3), 637–660.
- Woodrum, E. (1984). “Mainstreaming” Content Analysis in Social Science: Methodological Advantages, Obstacles, and Solutions. *Social Science Research*, 13(1), 1–19. doi:10.1016/0049-089X(84)90001-2
- Yu, C. H., Jannasch-Pennell, A., & Digangi, S. (2011). Compatibility Between Text Mining and Qualitative Research in the Perspectives of Grounded Theory , Content Analysis , and Reliability. *The Qualitative Report*, 16(3), 730–744.
- Yu, D., & Hang, C. (2010). A Reflective Review of Disruptive Innovation Theory. *Journal of Environmental Studies and Sciences*, 12(4), 435–452. doi:10.1111/j.1468-2370.2009.00272.x
- Zhang, Y., & Wildemuth, B. (2009). Qualitative Analysis of Content. In B. Wildemuth (Ed.), *Applications of Social Research Questions in Information and Library Science* (pp. 308–319). Greenwood Publishing Group Inc.