

This is the peer reviewed version of the following article:

Comunità virtuali di marca e innovazione di prodotto: Un modello di identificazione automatica dei Lead user basato su analisi testuale / Marchi, G.. - (2009). (Convegno SIM (Società Italiana Marketing) Firenze 2009).

*Terms of use:*

The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

12/08/2026 22:49

*Comunità virtuali di marca e innovazione di prodotto: un modello di classificazione degli user basato su analisi testuale*

**Stefano Bordoni\* – Gianluca Marchi\*\***

\*Dipartimento di Economia Politica, \*\*Dipartimento di Economia Aziendale  
Università di Modena e Reggio Emilia

**Abstract**

Il paper analizza la conoscenza del prodotto come fattore di classificazione e selezione dei membri più innovatori nelle comunità virtuali di marca. Con un approccio multi-metodo, il lavoro, che ha come ambito applicativo il blog Ducati Hypermotard, analizza i contenuti e i temi ricorrenti nei messaggi degli user e la loro distribuzione in relazione al grado di potenziale collaborativo del consumatore. La possibilità di stabilire un legame tra conoscenza dello user, linguaggio utilizzato e grado di *lead-userness* è confermata. Infine, è proposto un modello per la classificazione di testi postati sul blog in grado di assegnare automaticamente il profilo di *lead-userness* del mittente e di selezionare, pertanto, i commenti degli utenti più competenti in termini di prodotto e più innovativi.

**Parole chiave**

Innovazione di prodotto; comunità virtuale di marca; lead user; conoscenza di prodotto; analisi del contenuto; text-mining

**English abstract**

*(Virtual brand communities and product innovation: A users classification model based on text analysis)*

The paper analyses product knowledge as a key factor for classifying and selecting the more innovative users within virtual brand communities. Based on a multi-method approach, the work is applied to the Ducati Hypermotard blog, with the aim to study the recurrent contents and subjects observed in users' messages and their distribution in relation to the lead-user status of the members. The existence of a link between user's product knowledge, text contents, and lead user status is confirmed. Finally, a model for the automatic text-based classification of messages and users is proposed.

**Key Words**

Product innovation; virtual brand community; lead user; product knowledge; content analysis; text-mining

## 1. Introduzione

Il crescente interesse della letteratura e del management verso le comunità virtuali di marca è andato intrecciandosi, negli ultimi anni, col tema della conoscenza di mercato a supporto dei processi di innovazione di prodotto dell'impresa. Sulla scia dei primi contributi, l'attenzione specifica di questo lavoro è rivolta alla definizione di modelli che, a partire da popolazioni vaste ed eterogenee di partecipanti a comunità, classifichino e selezionino i consumatori potenzialmente più interessanti per capacità di collaborare all'innovazione. In particolare, ci si focalizza sulla conoscenza del prodotto come fattore di selezione dei membri più innovatori. La capacità del cliente di collaborare al processo d'innovazione di impresa dipende strettamente dalla conoscenza che egli possiede per generare nuove idee e soluzioni per il prodotto.

L'identificazione delle dotazioni cognitive dei partecipanti ad una comunità virtuale è compito molto complesso, data la difficoltà a sviluppare relazioni comunicative profonde e dirette con ognuno di essi. In effetti, la conoscenza nella Rete si manifesta principalmente attraverso testi e – come ben noto – il rapporto che lega conoscenza, linguaggio e testo è tutto tranne che lineare. Inoltre, i testi rilevanti non galleggiano sempre sulla superficie del web, ma si trovano spesso immersi in un'immensità di altro testo poco o per nulla rilevante per le decisioni di impresa. Per questo motivo, affrontare il problema della selezione dei testi più interessanti, della profilazione degli individui associati a quei testi e della capacità di catturare da essi conoscenza utile ai fini della decisione, rappresenta una tra le frontiere più ambiziose per la ricerca e la pratica di marketing applicate al Web2. La tesi del paper è che l'analisi del linguaggio utilizzato dai membri della comunità e contenuto nei loro messaggi testuali può costituire un terreno di ricerca proficuo per la comprensione dei diversi livelli di conoscenza di prodotto dei consumatori e, di conseguenza, del loro grado di *lead-userness*. I membri di una comunità, pertanto, possono essere utilmente differenziati in funzione del linguaggio utilizzato, per “le cose che dicono” e per “come le dicono”. Dopo aver trattato alcuni temi concettuali rilevanti per la comprensione del fenomeno dell'innovazione di prodotto nelle comunità di marca virtuali, il paper va ad approfondire il legame tra conoscenza e linguaggio degli user. La seconda sezione del lavoro si chiude con l'identificazione delle proposizioni di ricerca. Nella terza parte si delinea il disegno di ricerca, l'ambito di applicazione (il blog Hypermotard Ducati) e la metodologia utilizzata. La lettura dei risultati e la loro discussione, insieme alle principali implicazioni manageriali e ai possibili sviluppi futuri della ricerca, chiude il lavoro nelle sue due ultime sezioni.

## 2. Il framework teorico

### 2.1 Consumatore, innovazione e comunità

L'importanza della conoscenza di mercato per il processo di innovazione dei prodotti e per il vantaggio competitivo delle imprese è stata evidenziata a più riprese dalla

letteratura, in particolare negli approcci cognitivi all'innovazione (Verona, 2000). Nei contributi che pongono l'ambiente di mercato al centro del processo innovativo, i clienti, in particolare, sono visti come specifiche fonti di conoscenza da attivare in maniera diretta e interattiva. Questa attenzione è stata soprattutto forte negli studi della scuola nordica, nei servizi e nel marketing industriale, e in tutte quelle correnti di ricerca in cui il consumatore non è più il tradizionale soggetto passivo di analisi (Zaltman, 1997), ma è anche creatore di conoscenza e produttore di senso per l'innovazione. Sono soprattutto gli studi sul collaborative marketing (Cillo, 2004) ad enfatizzare il ruolo del consumatore come "co-creatore" del prodotto, attraverso l'interazione con l'impresa nel processo di innovazione. Soprattutto nell'ambito dei beni di consumo, il problema cruciale in queste impostazioni risiede nel trovare la logica per individuare livelli più aggregati di domanda omogenea a partire da singoli bisogni personali rilevati nell'interazione col cliente.

Al superamento dei limiti di un approccio atomistico alla collaborazione col mercato ha concorso l'introduzione più compiuta del concetto di comunità negli studi di marketing (Schouten e McAlexander, 1995; Cova, 1997). I consumatori raramente innovano in isolamento, «but rather in cooperation with like-minded people – acquaintances, colleagues, and friends» (Füller, Jawecki e Mühlbacher, 2007, p. 60). Le comunità sono luoghi dove generalmente si addensano consumatori con conoscenze di prodotto mediamente elevate e che possono arrecare nuova conoscenza utile all'impresa, anche in assenza di una diretta interazione e comunicazione con essa. La conoscenza degli individui si arricchisce durante il processo di socializzazione interno alla comunità, ampliandosi di valori simbolici e tribali. Nelle comunità di marca (Muniz e O'Guinn, 2001; Mc Alexander, Schouten e Koenig, 2002; Cucco e Dalli, 2008), in particolare, i consumatori volontariamente si incontrano per scambiare e condividere informazioni su una certa marca e per discutere sui nuovi prodotti ed il loro miglioramento (Kozinets, 2002).

## 2.2 Innovazione e comunità virtuali di marca

Un particolare interesse è stato rivolto alle comunità di marca operanti in ambiente virtuale (McWilliam, 2000). L'avvento delle nuove tecnologie di rete ha infatti facilitato il coinvolgimento dei consumatori in relazioni collaborative multidirezionali, asincrone e a distanza con l'impresa. È questo il tema principale affrontato dal filone della *collaborative innovation* (Prandelli e Verona, 2006). Le comunità virtuali sono rappresentabili come grandi serbatoi di conoscenza relativa al prodotto (comunità di informazione), che in determinate circostanze possono diventare importanti fonti di innovazione (Sawhney e Prandelli, 2000; Sawhney, Prandelli e Verona, 2005), fino a trasformarsi in vere e proprie comunità di innovazione (Von Hippel, 2005).

Le comunità di marca, indipendentemente dalla loro natura spontanea o programmata, solo in alcuni casi nascono come aggregazioni specificamente dedicate all'innovazione di prodotto. Quando le ragioni alla base della loro genesi sono di natura più generale, è normale osservare come i membri non sono tutti allo stesso

modo utili per le decisioni di impresa. Anche quando, poi, le comunità si formano su specifici problemi di innovazione, il contributo dei diversi membri risulta piuttosto eterogeneo. Aldilà dell'enfasi posta da molti studi sulla natura collettiva dell'innovazione e sulla matrice collaborativa del processo di creazione di nuova conoscenza (Sawhney e Prandelli, 2000), si è convinti che molto di questo processo di genesi distribuita delle idee resti ad oggi inesplorato. In realtà, non tutte le idee e i suggerimenti che giungono dagli user sono ugualmente interessanti per il successo dei nuovi prodotti. Ben lungi da essere un luogo dove l'interazione avviene tra pari, dentro alla comunità solo un piccolo gruppo di membri generalmente possiede «a set of distinctive characteristics differentiating them from the majority of users» (Piller e Walcher, 2006, p. 309). Solo se si comprendono bene le differenze che esistono nei contributi innovativi dei singoli partecipanti, possono emergere i soggetti che più di altri riescono a “dire la loro” e generare nuova conoscenza, che hanno competenze e motivazioni per farlo, che sono legittimati a dare senso alla conoscenza che emerge, valutare le complementarità ed integrarla con altra conoscenza esistente.

### *2.3 Lead user virtuali e conoscenza di prodotto*

In forza di questa prospettiva, assume crescente importanza formulare e testare modalità nuove di classificazione e selezione dei membri delle comunità on-line, in grado di tenere conto della specificità del contesto e coerenti con le finalità conoscitive dell'impresa (Kozinets, 1999). Una migliore selezione supporta la possibilità di coinvolgere i clienti innovatori in team di sviluppo di nuovi prodotti e nelle altre attività utili alla messa a punto e al lancio di nuovi prodotti. Aggiungiamo anche che questi modelli di classificazione e selezione devono rispondere ad un ulteriore requisito fondamentale: essere utilizzabili dal management con relativa facilità, implementati con costi non eccessivi rispetto ai risultati attesi e, soprattutto, coerenti con i tempi dei processi decisionali.

La proposta di criteri di classificazione adeguati passa per la formulazione di un profilo di cliente innovatore in grado di contribuire al processo di sviluppo di nuovi prodotti in un ambiente virtuale. Il concetto di cliente innovatore che nella letteratura ha trovato il maggior grado di consenso è quello di *lead user* (Von Hippel, 1986; Urban, Von Hippel, 1988), osservato, nei contributi più recenti, anche all'interno di ambienti web (Franke e Shah, 2003; Füller, Jawecki e Mühlbacher, 2007). I *lead user* sono quei clienti che, fortemente incentivati a partecipare al processo di innovazione per il perseguimento di un beneficio economico e non, grazie all'esperienza pratica nell'uso del prodotto e alla capacità di analisi prospettica, riescono a rappresentare il bisogno di un prodotto innovativo in anticipo rispetto al resto del mercato. Insieme alle motivazioni, sono dunque le capacità ad essere al centro del modello di rappresentazione del *lead user* (Lüthje e Herstatt, 2004).

È, in particolare, la caratteristica della conoscenza di prodotto ad apparire importante per predire la capacità di innovazione dello user. Solo i consumatori che hanno «the real-world experience needed to problem solve and provide accurate data to inquiring market researchers» (Von Hippel, 1986, p. 791) possono ambire a contribuire

all'innovazione di prodotto. Essi «do not have to imagine themselves in a yet non-existing situation [...] users facing new needs will find it easier to make sense of innovation-related information because it fits with their cognitive structure [...] These users are therefore more likely to recognize, to memorize and to apply pieces of information that may be relevant for developing solutions to their needs» (Lüthje e Herstatt, 2004, pp. 557-558).

Il presente contributo si focalizza proprio sulla conoscenza del prodotto come fattore principale di classificazione e selezione dei *lead user*. È stato, infatti, osservato che i partecipanti ad una comunità sono molto eterogenei tra loro in termini di dotazione di conoscenza di prodotto (Nambisan, 2002). Specialmente nei mercati di beni di consumo, il consumatore 'medio' spesso non dispone di conoscenza sufficiente per la piena esplicitazione degli attributi di prodotto in grado di soddisfare i propri bisogni. Questo vale sia per la fase di generazione delle idee che per quella di sviluppo del prodotto, quella cioè che riguarda la trasformazione del prototipo in prodotto finale. Lettl (2007) evidenzia, ad esempio, come, nella fase di sviluppo di un'innovazione radicale di prodotto, i contributi di tipo più attivo sono dati da user dotati di elevate competenze nel loro specifico dominio cognitivo di utilizzatori. Essi devono saper sviluppare una profonda comprensione degli elementi rilevanti nel dominio, gestire l'incertezza relativa ai benefici che il proprio contributo potrebbe apportare al prodotto finale e possedere un elevato know how tecnologico relativamente all'utilizzo del prodotto. Questa ultima caratteristica, in particolare, è essenziale, perché consente agli user innovatori di valutare immediatamente la fattibilità tecnica delle soluzioni di sviluppo proposte. Anche quando l'innovazione attesa è più di tipo incrementale, la capacità del cliente di collaborare al processo d'innovazione di impresa e generare nuove idee per il prodotto (*lead-userness status*) dipende strettamente dalla sua conoscenza.

#### 2.4 Conoscenza e linguaggio nelle comunità virtuali di marca

L'identificazione delle dotazioni cognitive dei partecipanti ad una comunità virtuale è, tuttavia, compito molto complesso, data la difficoltà a sviluppare relazioni comunicative profonde e dirette con ognuno di essi, limitante sovente alla sola raccolta di variabili anagrafiche (attraverso *form* o questionari on-line). Inoltre, la natura prevalentemente testuale degli artefatti attraverso cui si manifesta la presenza dell'individuo nelle comunità virtuali di marca<sup>1</sup>, impone di tenere conto di alcuni specifici adattamenti rispetto alle modalità di classificazione e selezione dei *lead user* utilizzati in ambienti non virtuali. Infine – e qui sta il limite più profondo – ben difficilmente ad un testo può essere associato un significato preciso ed univoco indipendentemente dal contesto culturale ad esperienziale in cui l'individuo è calato, contesto normalmente indagato attraverso tecniche di analisi mutuata dalla tradizione antropologica o dall'ermeneutica (Thompson, 1997; Sitz, 2008).

---

<sup>1</sup> Per una classificazione ampia dei *tools*, non solo testuali, impiegabili nei processi di innovazione collaborativa sul web, si rimanda a Prandelli, Verona e Raccagni (2006).

Pur consapevoli di queste criticità, l'analisi del linguaggio utilizzato dai membri della comunità e contenuto nei loro messaggi testuali può costituire un terreno di ricerca proficuo per la comprensione dei diversi livelli di conoscenza di prodotto dei consumatori e, di conseguenza, del loro grado di *lead userness*. La tesi del lavoro, pertanto, è che i membri di una comunità di marca possono essere differenziati in funzione del linguaggio utilizzato, per "le cose che dicono" e per "come le dicono". Nonostante la rilevanza del concetto di conoscenza di prodotto e la natura prevalentemente testuale della comunicazione in ambiente virtuale, la letteratura sull'innovazione di prodotto ha dedicato poca attenzione al legame tra conoscenza di prodotto e linguaggio degli user. In parte, questo è certamente dovuto alla già accennata difficoltà nella definizione teorica e nella rappresentazione operativa di tale legame. Non tutta la conoscenza, infatti, è articolabile, separabile dal contesto di pratica ed esprimibile linguisticamente (Winter, 1987). Inoltre, il trasferimento della conoscenza presuppone la condivisione collettiva di un linguaggio tra fonti e destinatari e, spesso, anche una significativa condivisione delle pratiche di uso del prodotto, poiché è solo l'ancoraggio ad una comune base esperienziale che permette ai membri di condividere una relazione univoca tra significante e significato (Hislop, 2005).

Da questo punto di vista, la comunità di marca on-line, come luogo specifico di produzione e diffusione di conoscenza, presenta una caratteristica interessante: è un contesto dove pratiche esperienziali simili sono abbastanza diffuse. Questo consente l'instaurarsi di una relazione stretta tra pratica e linguaggio anche nel contesto collettivo. Siamo, quindi, di fronte ad un sistema di relazioni comunicative piuttosto coeso sul piano delle pratiche e, conseguentemente, della condivisione dei linguaggi.

## 2.5 Le domande di ricerca

Nonostante il contesto sia favorevole al trasferimento di conoscenza, anche nelle comunità on-line non tutti i membri mostrano competenze simili nel processo di comunicazione. L'eterogeneità tra membri permane. Questo lavoro vuole pertanto indagare se l'analisi del linguaggio e l'utilizzo di modelli *text-based* di classificazione e selezione possa essere d'aiuto nella comprensione sia del diverso grado di conoscenza del prodotto dei partecipanti che della loro differente volontà di partecipazione al processo innovativo (per frequenza ed entità dei contributi, temi trattati, grado di approfondimento degli argomenti).

Il primo aspetto da considerare parte dalla valutazione che la comunità di marca on-line è un luogo dove i processi di comunicazione assumono intenzionalmente una valenza sociale (Bagozzi e Dholakia, 2002). Come dice Kozinets (2002), «the posting of computer text is a social action (a communicative act or 'language game')» (p. 64). In una comunità, attraverso la produzione del testo, si estrinseca un'azione comunicativa intenzionale da parte del consumatore partecipante (Bara, 1999). Il testo è raccolto ed interpretato da una specifica audience, alla quale appartengono individui caratterizzati da un forte attaccamento alla marca e intenzionati a scambiare opinioni ed esperienze d'uso sulla marca e i prodotti correlati. Sotto questo profilo, la comunità

di marca è un contesto che stimola gli individui, soprattutto quelli più esperti, a comunicare e ad esplicitare la loro esperienza, codificandola in forma di linguaggio per metterla a disposizione di tutti. I membri esperti “si sentono” sollecitati a svolgere un compito comunicativo, mossi o da un senso di “obbligazione” verso la community (Muniz e O’Guinn, 2001) o dalla ricerca di forme di para-socialità legate alla dimensione della *self-exhibition* (Cova e Pace, 2006). Inoltre, anche lo sforzo comunicativo richiesto al consumatore appare diverso: maggiore è il suo livello di expertise, minore lo sforzo che dovrà compiere (Alba e Hutchinson, 1987). La capacità di trasferire conoscenza varia, quindi, al variare del grado di familiarità col prodotto dei membri. Quindi, si può presumere che, all’interno della comunità virtuale, i compiti più complessi in termini di produzione del messaggio, sia sul piano quantitativo che qualitativo, vengano compiuti da persone esperte<sup>2</sup>.

Prop.1: Maggiore è il livello di *lead-userness* di un individuo, maggiore è la conoscenza del prodotto posseduta, maggiore la sua legittimazione e capacità di comunicare all’interno della comunità, maggiore è l’entità dei contributi offerti.

Un secondo aspetto da esaminare riguarda la tipologia prevalente di contenuti osservabili nei messaggi testuali ed il loro legame con la qualità del contributo all’innovazione del singolo user. Rispetto a questo, si possono attendere situazioni molto diverse in funzione del tipo di prodotto e della specificità dei settori. Di Maria e Finotto (2008) argomentano, ad esempio, che nei settori del Made in Italy le componenti estetiche e di comunicazione prevalgono nettamente su quelle tecniche in sede di innovazione di prodotto. Generalmente, però, è la conoscenza a carattere tecnologico ad essere più correlata alla propensione a dare supporto all’innovazione (Lettl, 2007). «Technical knowledge consists of know-how concerning the product architecture, the used materials, and the applied technologies in a product category. [...] User firms or end users need to have this understanding if they want to translate the wants and needs into specific product specifications» (Lüthje e Herstatt, 2004, p. 560). Dalla distribuzione eterogenea di conoscenza tecnica può derivare quindi una distribuzione disomogenea anche dei contenuti trattati nei messaggi, a seconda del grado di *lead-userness* dell’individuo.

Prop. 2: Maggiore è il livello di *lead-userness* di un individuo, maggiore è la presenza nei suoi messaggi di contenuti a carattere tecnico.

Un ulteriore tratto di specificità, che connota il legame tra linguaggio e *lead-userness*, riguarda il grado di soddisfazione nei confronti dei prodotti esistenti dell’impresa. In letteratura è stato evidenziato come, tra gli indicatori che si possono utilizzare per identificare i *lead user*, oltre alla capacità di anticipare i bisogni (collegata alla conoscenza esperienziale di prodotto), vi sia anche l’insoddisfazione verso i prodotti esistenti (Lüthje e Herstatt, 2004) o in fase di sviluppo, che spinge i consumatori con

---

<sup>2</sup> Le ragioni che spingono ad intervenire con maggiore intensità e qualità in una comunità virtuale di marca sono ovviamente molteplici. Füller, Jawecki e Mühlbacher (2006), ad esempio, nel loro studio sulle comunità di innovazione per le scarpe da pallacanestro, hanno evidenziato come i contributi più rilevanti all’innovazione di prodotto provengono da individui mossi, più che da specifici bisogni da soddisfare (solo una minoranza), dal divertimento in sé che traggono dal pensare nuovi prodotti e suggerire modifiche e cambiamenti.

un grado elevato di coinvolgimento verso la marca a fare emergere pubblicamente le criticità principali.

Prop. 3: Maggiore è il livello di *lead-userness* di un individuo, maggiore l'intensità di presenza di giudizi critici o anche negativi nei suoi messaggi.

### 3. Disegno di ricerca e metodologia

Il disegno di ricerca si fonda su una logica di esperimento quasi-naturale condotto attraverso l'analisi longitudinale di una comunità virtuale di marca nel settore motociclistico, il blog Hypermotard Ducati, istituito per supportare lo sviluppo e il lancio di un nuovo modello di prodotto.

In una prima fase di ricerca, la comunità è stata attentamente osservata adottando tecniche netnografiche (Kozinets, 2002) basate sull'osservazione del linguaggio testuale contenuto nei messaggi dei partecipanti al blog. Sono stati individuati 2.403 messaggi inviati al blog Ducati nel periodo compreso tra il marzo 2006 e il maggio 2007<sup>3</sup>. A partire da questo primo dataset, i 2.403 messaggi sono stati accorpati in relazione al nome del mittente, ottenendo in questo modo un secondo dataset *Utenti membri* con 824 record, pari cioè al numero dei membri che hanno almeno una volta partecipato al blog anche in modo del tutto occasionale. Su questo dataset, attraverso un'analisi del contenuto basata sull'operato di due *rater* formati da uno dei due autori<sup>4</sup>, sono state analizzate la coerenza dei messaggi rispetto ai temi lanciati dal moderatore, il numero di messaggi inviati e l'assiduità della presenza nelle discussioni (*commitment*), la conoscenza dei blogger rispetto al prodotto (*product knowledge*), il grado di allineamento dei suggerimenti rispetto all'identità di marca come proxy della capacità previsiva (*forecasting ability*). L'analisi ha permesso di assegnare un punteggio di *lead-userness* compreso tra 0 e 4 a tutti i partecipanti e di individuare così quei soggetti più importanti per l'attività di innovazione (Kozinets, 1999)<sup>5</sup>.

A partire dai risultati di questa prima fase, la ricerca è stata indirizzata lungo due percorsi metodologici di ricerca.

1) Un primo obiettivo di approfondimento, tipico della *content analysis* in senso stretto (Neuendorf, 2002), ha provato a capire la frequenza con cui ricorrono alcuni

---

<sup>3</sup> Per un'analisi più approfondita di questa prima parte della ricerca si rimanda a Marchi e Etzi (2007).

<sup>4</sup> Per saggiare l'affidabilità del loro operato, ai due *rater* è stato affidato il compito di svolgere l'analisi del contenuto sulla stessa intera collezione di testo ma in due momenti separati nel tempo (a sei mesi di distanza). In una fase successiva, le valutazioni divergenti sono state analizzate congiuntamente dai due *rater*, comprese le ragioni che avevano portato alla difformità di giudizio e, infine, rese uniformi.

<sup>5</sup> La misurazione utilizzata tiene conto del carattere continuo della variabile di *lead-userness* o '*leading edge status*' per usare la definizione di Morrison *et al.* (2004). L'unimodalità di tale variabile induce a ritenere che una rappresentazione dicotomica e bivariata del concetto di *lead user* rischi di disperdere molta informazione utile.

temi considerati significativi rispetto all'insieme dei messaggi inviati. Tra i possibili diversi approcci di *mining* al corpus di documenti contenuti nel blog, l'analisi del contenuto testuale è stata condotta con un software di *content analysis rule-based*<sup>6</sup>. Operativamente, l'approccio ha previsto la costituzione di un dizionario personalizzato composto da circa 2.000 *keyword* considerate rilevanti in questo contesto di analisi<sup>7</sup>. Si è, poi, proceduto ad una categorizzazione dei contenuti in grado di tenere conto delle specifiche domande di ricerca formulate. Sono state contemplate le aree tematiche entro cui tendono ad aggregarsi i principali attributi di prodotto, cercando di considerare sia le dimensioni tecniche e tangibili che quelle soft e immateriali dell'offerta (Brand, Componenti tecniche, Processo di comunicazione, Contesti di utilizzo, Performance,...). Nella predisposizione del piano di categorie si è, quindi, tenuto conto della possibilità di formulare giudizi e commenti, positivi o negativi, sul nuovo prodotto oggetto del blog. Alla fine, sono 24 le categorie lessicali con cui si è organizzato e raggruppato il testo estratto ad un livello logico-semanticamente più elevato (Tab. 1)<sup>8</sup>.

*Tab. 1 – Alcune categorie lessicali di contenuto utilizzate*

| <i>Categorie</i> | <i>Esempi di keywords</i>                                                             |
|------------------|---------------------------------------------------------------------------------------|
| BRAND            | Immagine, made, sport, storia...                                                      |
| COMPONENTI       | coppia, forcella, frizione, motore, ruote, serbatoio, sospensioni, testa, valvole,... |
| COMUNICAZIONE    | foto, lancio, marketing, moda, nome, pubblicità, style...                             |
| DISTRIBUZIONE    | concessionario, consegna, mercato...                                                  |
| IDENTITÀ         | amici, anima, appassionati, ducalista, forum, idee, passione...                       |
| LUOGHI           | aria, bar, curve, gara, giro, montagna, tornanti...                                   |
| MARCHE MOTO      | aprilvia, bmw, giapponesi, honda...                                                   |
| MODELLI DUCATI   | monster, multistrada, pompone...                                                      |
| PREZZO           | prezzo, soldi...                                                                      |
| PRODOTTO         | cavalli, cilindrata, colore, design, gamma, italiano, linea, peso, versioni...        |

*Fonte: nostra elaborazione*

Col supporto di tecniche di analisi multivariata, tra cui l'Analisi delle Corrispondenze (Greenacre, 1984), si è cercato di capire, poi, se i gruppi di utenti classificati con

<sup>6</sup> QDAMiner, Wordstat, Provalis Research. Vedi <http://www.provalisresearch.com/>.

<sup>7</sup> Ogni *keyword* può corrispondere ad una singola parola (es. DUCATISTA), ad una frase (es. NEI\_PROSSIMI\_TRE\_ANNI) o a forme lessicali derivate o composte (es. ANELLO\_DI\_FERMO\_SEEGER).

<sup>8</sup> All'interno del dizionario è contenuto anche un insieme di regole per identificare ed estrarre dai messaggi le porzioni di testo considerate rilevanti, così da potere ricavare le *keyword* in relazione ad un dato contesto e ad uno specifico significato minimizzando gli errori di estrazione ed il numero di falsi positivi (*part of speech disambiguation*). Le regole impongono delle condizioni di estrazione collegando i contenuti delle diverse categorie attraverso operatori booleani e di prossimità, come, ad esempio, nel caso: "NEG\_COMP [#COMP NEAR #NEGATIVO & #COMP NOT NEAR #NEGAZIONI /C 5]". Questa regola andrà ad estrarre la frase "la forcella (categoria COMPONENTI TECNICHE) mi ha dato dei problemi (categoria NEGATIVO)", ignorando invece la frase "la forcella (categoria COMPONENTI TECNICHE) non (categoria NEGAZIONI) mi ha dato dei problemi (categoria NEGATIVO)".

crescenti valori di *lead-userness* esprimano in proporzione quote diverse (crescenti) di testo ed idee significative. Questo ha permesso di osservare i temi più discussi all'interno di ciascun gruppo e messo in evidenza le differenze distributive nel tipo e nella significatività dei testi inviati. In particolare, l'analisi è stata finalizzata a verificare la presenza di argomenti specifici, quali suggerimenti, giudizi e confronti con i prodotti concorrenti, che potessero caratterizzare in modo significativo i messaggi dei consumatori innovatori rispetto agli altri partecipanti al blog.

2) Un secondo obiettivo, tipico delle tecniche di *text-mining* (Dulli, Polpettini e Trotta, 2004; Zanasi, 2005; Bordoni, 2008), ha previsto la costruzione di un modello di classificazione dell'utente del blog e/o del messaggio postato in grado di assegnare automaticamente il grado di *lead-userness* del mittente e di selezionare, pertanto, i commenti degli utenti più competenti in termini di prodotto. La classificazione automatica del testo è una tecnica di raggruppamento supervisionata che permette di assegnare a ciascun record (caso) un valore stimato della classe di appartenenza rispetto ad una variabile dipendente. In questo contesto è stato svolto un apprendimento induttivo sul dataset (Sebastiani, 1999) *Utenti membri* contenente la classificazione manuale di 49 clienti caratterizzati da un grado positivo di *lead-userness* (Marchi e Etzi, 2007). Ciò ha permesso di addestrare il modello di classificazione confrontando il contenuto del testo di ciascun utente con il corrispondente valore della variabile target già inserita nella prima fase della ricerca. La maggiore o minore ricorrenza nel testo di una o più *keyword* con i diversi valori di *lead-userness* permette al modello di associare il linguaggio utilizzato ad uno specifico valore risultato, in questo caso dato dal punteggio di *lead-userness* compreso tra 0 e 4. Al termine dell'addestramento, il modello costruisce una matrice di confusione, che riassume i risultati rispetto al dataset di training, confrontando i valori reali di ciascuna classe (righe della matrice) con quelli calcolati (colonne).

#### 4. Risultati

L'analisi del contenuto (su giudizi, suggerimenti, confronti) restituisce una prima evidenza sulla natura dei blogger, che aiuta a comprendere meglio il contesto di rilevazione (Tab. 2).

Osservando gli elevati valori di frequenza dei suggerimenti espressi, sia rispetto alle dimensioni tecniche che non tecniche dell'offerta Ducati, si può notare come sia elevato il senso di appartenenza e il coinvolgimento nei processi innovativi e creativi dell'azienda. Inoltre, i commenti di carattere positivo dei partecipanti superano ampiamente quelli negativi (93,6% del totale), in linea con quanto atteso in un contesto di blog dove la presenza di brand *passionate* è molto elevata e dove alto è il senso di appartenenza e il coinvolgimento nei processi innovativi e creativi dell'azienda.

L'analisi multivariata consente di affermare che agli utenti con maggior Grado di *Lead-Userness* (GLU) (da 0 a 4, rispettivamente 775, 16, 14, 12, 7 utenti) si associano

quote maggiori di testo significativo, cioè di testo riferibile ad una delle categorie lessicali considerate rilevanti ai fini dell'analisi (Tab. 3).

*Tab. 2 – Frequenza delle categorie di contenuto (parti di testo e messaggi).*

| <i>Categorie di contenuto</i>         | <i>Parti di testo significativo<br/>N. (%)</i> | <i>Messaggi con testi significativi<br/>N.</i> |
|---------------------------------------|------------------------------------------------|------------------------------------------------|
| Giudizi positivi_Ducati               | 342 (21,3%)                                    | 195                                            |
| Giudizi positivi_Prodotto             | 254 (15,8%)                                    | 160                                            |
| Giudizi positivi_Moto                 | 199 (12,4%)                                    | 137                                            |
| Termini Glossario Ducati              | 168 (10,5%)                                    | 102                                            |
| Giudizi positivi_Componenti           | 155 (9,7%)                                     | 88                                             |
| Consigli_Prodotto                     | 82 (5,1%)                                      | 59                                             |
| Giudizi positivi_Ducati vs competitor | 58 (3,6%)                                      | 48                                             |
| Giudizi positivi_Luoghi d'utilizzo    | 58 (3,6%)                                      | 41                                             |
| Giudizi positivi_Brand                | 54 (3,4%)                                      | 47                                             |
| Consigli_Comunicazione                | 54 (3,4%)                                      | 43                                             |
| Giudizi positivi_da Glossario         | 50 (3,1%)                                      | 41                                             |
| Consigli_Componenti                   | 50 (3,1%)                                      | 39                                             |
| Giudizi negativi_Prodotto             | 15 (0,9%)                                      | 11                                             |
| Giudizi negativi_Prezzo               | 12 (0,7%)                                      | 11                                             |
| Giudizi negativi_Moto                 | 12 (0,7%)                                      | 10                                             |
| Giudizi negativi_Ducati               | 9 (0,6%)                                       | 8                                              |
| Giudizi negativi_Distribuzione        | 7 (0,4%)                                       | 7                                              |
| Giudizi negativi_da Glossario         | 6 (0,4%)                                       | 4                                              |
| Giudizi negativi_Componenti           | 5 (0,3%)                                       | 3                                              |
| Giudizi negativi_Luoghi d'utilizzo    | 4 (0,2%)                                       | 4                                              |
| Confronti a favore                    | 4 (0,2%)                                       | 4                                              |
| Giudizi negativi_Ducati vs competitor | 4 (0,2%)                                       | 3                                              |
| Giudizi negativi_Brand                | 3 (0,2%)                                       | 3                                              |
| Giudizi negativi_Prezzo               | 1 (0,1%)                                       | 1                                              |
| Totale                                | 1.606 (100,0%)                                 | 1.069                                          |

*Fonte: nostra elaborazione*

*Tab. 3 – Distribuzione degli utenti per significatività di testo.*

|                                  | GLU 0 | GLU 1  | GLU 2 | GLU 3 | GLU 4  |
|----------------------------------|-------|--------|-------|-------|--------|
| Numero utenti                    | 775   | 16     | 14    | 12    | 7      |
| % utenti con testo significativo | 47,1% | 100,0% | 85,7% | 91,7% | 100,0% |

*Fonte: nostra elaborazione*

La percentuale degli utenti di ciascuna tipologia di GLU (1, 2, 3, 4) che ha espresso un testo significativo rispetto ad una certa categoria di contenuto è superiore a quella degli altri utenti complessivamente e per 21 delle 24 categorie semantiche considerate (Tab. 4)<sup>9</sup>. All'interno della categoria GLU = 4, ad esempio, il 100% degli utenti esprime idee considerate significative in termini di giudizi positivi sulle componenti tecniche, sul prodotto e sui modelli Ducati. I risultati, quindi, sembrano confermare la

<sup>9</sup> Anche aggregando gli utenti con GLU > 0 in un'unica categoria il contenuto dei messaggi risulta comunque più espressivo ed efficace.

rilevanza della prima proposizione, suggerendo l'utilità, per il futuro, di formulare in modo più stringente ipotesi di ricerca ad essa ricollegabili.

*Tab. 4 - Tipologie di contenuto per grado di lead-userness (GLU). (% di testo significativo)*

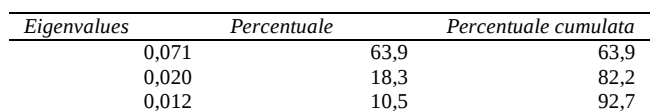
| <i>Categorie di contenuto</i> | <i>GLU 0</i> | <i>GLU 1</i> | <i>GLU 2</i> | <i>GLU 3</i> | <i>GLU 4</i> | <i>Chi2</i> | <i>P (2-tails)</i> |
|-------------------------------|--------------|--------------|--------------|--------------|--------------|-------------|--------------------|
| Giud positivi_Ducati          | 20,3%        | 68,8%        | 78,6%        | 75,0%        | 100,0%       | 86,4        | 0,000              |
| Giud positivi_Prodotto        | 17,5%        | 25,0%        | 57,1%        | 41,7%        | 100,0%       | 47,6        | 0,000              |
| Giud positivi_Moto            | 14,6%        | 31,3%        | 64,3%        | 33,3%        | 85,7%        | 54,3        | 0,000              |
| Termini Glos. Ducati          | 10,5%        | 6,3%         | 50,0%        | 58,3%        | 85,7%        | 79,6        | 0,000              |
| Giud posit_Componenti         | 9,0%         | 25,0%        | 35,7%        | 16,7%        | 100,0%       | 73,8        | 0,000              |
| Consigli_Prodotto             | 5,8%         | 0,0%         | 42,9%        | 25,0%        | 71,4%        | 79,4        | 0,000              |
| Giud pos_Duc. vs comp.        | 5,0%         | 0,0%         | 21,4%        | 8,3%         | 71,4%        | 63,1        | 0,000              |
| Giud positivi_Luoghi          | 3,7%         | 12,5%        | 28,6%        | 25,0%        | 42,9%        | 52,3        | 0,000              |
| Giud positivi_Brand           | 4,6%         | 12,5%        | 28,6%        | 16,7%        | 42,9%        | 37,2        | 0,000              |
| Consigli_Comunicazione        | 4,3%         | 0,0%         | 28,6%        | 33,3%        | 28,6%        | 44,7        | 0,000              |
| Giud posit_da Glossario       | 3,7%         | 18,8%        | 28,6%        | 16,7%        | 42,9%        | 50,1        | 0,000              |
| Consigli_Componenti           | 4,1%         | 6,3%         | 21,4%        | 8,3%         | 28,6%        | 18,5        | 0,002              |
| Giud negativi_Prodotto        | 0,8%         | 0,0%         | 14,3%        | 8,3%         | 28,6%        | 63,8        | 0,000              |
| Giud positivi_Prezzo          | 1,4%         | 0,0%         | 0,0%         | 0,0%         | 0,0%         | 0,7         | 0,983              |
| Giud negativi_Moto            | 0,9%         | 0,0%         | 7,1%         | 8,3%         | 14,3%        | 20,0        | 0,001              |
| Giud negativi_Ducati          | 0,6%         | 0,0%         | 0,0%         | 8,3%         | 28,6%        | 63,4        | 0,000              |
| Giud pos_Distribuzione        | 0,5%         | 0,0%         | 7,1%         | 8,3%         | 14,3%        | 30,7        | 0,000              |
| Giud negativi_da Glos         | 0,3%         | 0,0%         | 7,1%         | 8,3%         | 0,0%         | 29,1        | 0,000              |
| Giud negat_Componenti         | 0,4%         | 0,0%         | 0,0%         | 0,0%         | 0,0%         | 0,2         | 0,999              |
| Giud negativi_Luoghi          | 0,3%         | 0,0%         | 0,0%         | 0,0%         | 28,6%        | 115,3       | 0,000              |
| Confronti a favore            | 0,4%         | 0,0%         | 7,1%         | 0,0%         | 0,0%         | 13,2        | 0,022              |
| Giud neg_Duc. vs comp         | 0,3%         | 0,0%         | 7,1%         | 0,0%         | 0,0%         | 18,1        | 0,003              |
| Giud negativi_Brand           | 0,3%         | 0,0%         | 7,1%         | 0,0%         | 0,0%         | 18,1        | 0,003              |
| Giud negativi_Prezzo          | 0,1%         | 0,0%         | 0,0%         | 0,0%         | 0,0%         | 0,1         | 1,000              |

*Fonte: nostra elaborazione*

I risultati mostrano anche come i consumatori con  $GLU > 0$  evidenzino una maggiore predisposizione a strutturare i loro interventi testuali come consigli e osservazioni su alcuni specifici temi, come i contesti di utilizzo (luoghi), le moto, i prodotti Ducati e il brand. Più specificamente, in tutte le 6 categorie semantiche a contenuto prettamente tecnico (riferite a prodotti e a componenti della moto) i membri con  $GLU > 0$  evidenziano un numero di testi significativi statisticamente superiore. Nelle categorie di contenuto non direttamente riferibili ad elementi tecnici di valutazione (prestazioni generiche della moto, contesti di utilizzo, comunicazione, brand, concorrenti, distribuzione) la percentuale di soggetti con  $GLU > 0$  resta superiore, con la sola eccezione dei giudizi sui prezzi e dei giudizi negativi sulle componenti della moto. I risultati corroborano, quindi, anche la tenuta della seconda proposizione avanzata. I *lead user* tendono a inserire contenuti tecnici nei loro messaggi in modo notevolmente superiore rispetto agli altri partecipanti.

L'Analisi delle Corrispondenze (Graf. 1) ci fa osservare graficamente come, presso gli utenti non *lead user*, emerga la tendenza a parlare di prezzi o di aspetti più generici del mondo dei motori, mentre i gruppi di utenti con  $GLU > 0$  sono più prodighi di

*Graf. 1 – Analisi delle corrispondenze: Distribuzione delle categorie di contenuto per gruppi di user.*



In particolare, gli utenti col maggior grado di *lead-userness* (GLU = 4) fanno più numerose valutazioni critiche, soprattutto con riferimento ai contesti di utilizzo, ma anche al prodotto e alle sue prestazioni. I risultati confortano circa la sensatezza della terza proposizione avanzata, anche se il rinvio ad ulteriori ricerche sul tema appare necessario.

13

Da un punto di vista operativo, il modello può essere applicato a un dataset di record nuovi non classificati per individuare i *lead user* di altre community, senza richiedere ulteriori interventi manuali, con una precisione attesa solitamente inferiore a quella di training<sup>10</sup>, ma realisticamente vicina a quella calcolata sui dati di addestramento quanto più i dataset sono simili per dimensioni, contenuto e contesto di applicazione. Il risultato conferma l'ipotesi che i clienti con le potenzialità più elevate in termini di *lead-userness* tendono generalmente a “dire cose diverse” rispetto ai membri meno caratterizzati. Pertanto, sulla base di queste sole differenze, un sistema addestrato di estrazione è in grado, con buona approssimazione, di classificare i nuovi messaggi entranti e di assegnarne il membro mittente in un predefinito gruppo-tipo di consumatori.

## 5. Implicazioni per il management e la ricerca

Anche se la conoscenza assume nella comunità una dimensione collettiva, sono sempre singoli individui a proporre idee, a segnalare le migliori di altri utenti, a fungere da connettori sociali all'interno della comunità, contribuendo a propagare certe idee e non altre, certe valutazioni di un fenomeno e non altre. Comprendere i diversi ruoli di questi individui all'interno delle comunità, classificarli e selezionarli sulla base del contenuto dei loro messaggi, e non solo dei loro profili anagrafici, è compito essenziale per il management, soprattutto per l'attività di innovazione.

Il presente contributo, che ha natura esplorativa, prova ad aprire la strada all'individuazione di nuovi sistemi di classificazione degli utenti basati su analisi testuali e, quindi, sul linguaggio da loro utilizzato nella comunità, con l'obiettivo di stimare i diversi livelli di conoscenza di prodotto dei consumatori e, di conseguenza, il loro grado di *lead-userness*. Poiché il monitoraggio continuo su base manuale di testi depositati sul web da centinaia o migliaia di utenti è attività gravosa per tempi e costi, questo lavoro ha provato a dimostrare come possa assumere molta importanza, per il management, il ricorso a modelli più evoluti di estrazione dei partecipanti, in grado, con tempi predefinibili e costi sufficientemente limitati, di scremare quei soggetti che, per le loro specifiche caratteristiche e competenze, per la capacità di condivisione della conoscenza posseduta e per la coerenza con i percorsi di evoluzione strategica e tecnologica delle imprese, possono dare un sostanziale contributo al processo di sviluppo dei prodotti.

Il modello qui testato va in questa direzione di intervento: si può applicare a un dataset di record (messaggi/utenti) nuovi non ancora classificati per individuare con procedura automatica i *lead user* della community, con una precisione attesa

---

<sup>10</sup> Il tipo di metodo di validazione incrociata (*Leave one out*) e di parametri utilizzati nell'applicazione dell'algoritmo di classificazione (*K-Nearest neighbour*), possono sovrastimare le capacità predittive del modello, dando luogo al fenomeno di “overfitting”. In questo caso il modello potrebbe risultare molto preciso rispetto alle caratteristiche specifiche del training set, ma sensibilmente meno performante rispetto a dati nuovi.

realisticamente vicina a quella di un rater esperto. La sua applicazione a comunità diverse da quella da cui ha tratto origine aiuterà a validarne meglio l'efficacia e ad introdurre progressivi aggiustamenti nelle logiche di addestramento. Nonostante ciò, si ritiene che la dimensione *context-specific* resti decisiva per la natura del modello. Per contesti d'applicazione via via più lontani da quello d'origine, ci si può aspettare che l'efficacia del modello di classificazione e selezione si attenui.

Molto resta da mettere a fuoco anche nella riflessione teorica su alcune delle tematiche trattate in questo lavoro. Prima di tutto, appare interessante capire come la diversità delle motivazioni alla base della partecipazione alla comunità possa influire su capacità e intenzioni collaborative degli user e, soprattutto, sulle loro modalità di comunicare e trasferire conoscenza attraverso produzione linguistica di testo. Ad esempio, capire se la dimensione solo economica dei ritorni collegati al *free-revealing* (Harhoff, Henkel e Von Hippel, 2003) possa essere sufficiente a spiegare impegno e rilevanza del contributo innovativo dello user; o se, piuttosto, tale dimensione non debba essere integrata dalla considerazione del peso che elementi di legittimazione sociale o di *self-exhibition* possono esercitare sul comportamento degli user; o se, ancora più radicalmente, sia la separazione stessa, sul piano concettuale, delle fasi di produzione e distribuzione del surplus da innovazione a dovere essere superata (Cova, Dalli, 2009), all'interno di modelli molto diversi di relazione tra impresa e consumatore.

Il secondo ordine di problemi di carattere teorico sta tutto dentro al tema del legame, non facilmente identificabile, né sul piano concettuale né pragmatico, tra conoscenza e linguaggio. Insomma cosa resta da dire alla tradizionale analisi del contenuto dopo la "svolta linguistica" di Wittgenstein? In che misura è praticabile, in altre parole, una logica di intervento metodologico che ancori a categorie lessicali univoche i diversi livelli logico-semantici presenti nel linguaggio testuale? In questo lavoro, parte del problema è stato ridotto in radice, a partire dalla scelta del particolare "laboratorio" di applicazione del modello, una comunità di marca coesa sul piano dei linguaggi e delle pratiche comuni, così da sostanziare ancoraggi abbastanza stabili tra significanti e significati. I risultati qui ottenuti fanno intravedere che i pattern di linguaggio conservano alcune regolarità e mantengono una rilevanza euristica sufficiente a discriminare profili diversi di utenti. Non vi è dubbio, tuttavia, che, in contesti meno coesi, alcune di queste regolarità possano affievolirsi, riportando al centro del problema teorico e manageriale il nodo critico del legame tra conoscenza e linguaggio.

## **Bibliografia**

- Alba, J W.; Hutchinson J. W. (1987), Dimensions of Consumer Expertise, *Journal of Consumer Research* 13, March: 411-454.
- Bagozzi R.P., Dholakia U.M. (2002), Intentional social action in virtual communities, *Journal of Interactive Marketing*, 16, 2: 2-21.

- Bara B. G. (1999), *Pragmatica Cognitiva. I processi mentali della comunicazione*, Bollati Boringhieri, Torino.
- Bordoni S. (2008), *Analisi intelligenti dei dati per applicazioni di CRM analitico*, Pitagora.
- Cillo P. (2004), *Innovazione e mercato*, Carocci, Roma.
- Cova B. (1997), Community and consumption: Towards a definition of the linking value of products and services, *European Journal of Marketing*, 31, 3-4: 297-316.
- Cova B., Pace S. (2006), Brand community of convenience products: new forms of customer empowerment – the case “my Nutella The Community”, *European Journal of Marketing*, 40, 9/10: 1087-1105.
- Cova B., Dallì D. (2009), Working consumers: the next step in marketing theory, *Marketing Theory*, 9,3:315-339.
- Cucco R., Dallì D. (2008), 500 wants you. Un caso di convergenza tra retro-marketing, cooperative innovation e community management, *Economia & Management*, 2: 53-72.
- Di Maria E., Finotto V. (2008), Communities of consumption and Made in Italy, *Industry and Innovation*, 15, 2: 179-197.
- Dulli S., Polpettini P., Trotta M. (2004), *Text mining: teoria ed applicazioni*, Franco Angeli, Milano.
- Franke N., Shah S. (2003), How communities support innovative activities: an exploration of assistance and sharing among end-users, *Research Policy*, 32: 157-178.
- Füller J., Jawecki G., Mühlbacher H. (2007), Innovation creation by on-line basketball communities, *Journal of Business Research*, 60: 60-71.
- Greenacre, M. (1984), *Theory and Applications of Correspondence Analysis*, Academic Press, Orlando, Florida.
- Harkoff D., Henkel J., Von Hippel E. (2003), Profiting from voluntary information spillovers: how users benefit from freely revealing their innovation, *Research Policy*, 32: 1753-1769.
- Hislop D. (2005), *Knowledge management in organizations*, Oxford University Press, Oxford.
- Kozinets R. (1999), “E-Tribalized marketing?: The strategic implications of virtual communities of consumption”, *European Management Journal*, 13, 3: 252-264.
- Kozinets R. (2002), The field behind the screen: using netnography for marketing research in on-line communications, *Journal of Marketing Research*, 39,1: 61-72.
- Lettl C. (2007), User involvement competence for radical innovation, *Journal of Engineering and Technology Management*, 24: 53-75.
- Lüthje C., Herstatt C. (2004) The Lead User method: an outline of empirical findings and issues for future research, *R&D Management*, 34, 5: 553-568.
- Marchi G., Etzi A. (2008), Comunità virtuali e collaborative innovation: un modello per classificare e selezionare i lead user, *Proceedings del VII Congresso Internazionale “Le tendenze del marketing”*, Venezia, Gennaio 17-19.

- McAlexander J.H., Schouten J.W., Koenig H.F. (2002), Building brand community, *Journal of Marketing*, 66: 38-54.
- McWilliam G. (2000), Building stronger brands through on-line communities, *Sloan Management Review*, Spring: 43-54.
- Morrison P.D., Roberts J.H., Midgley D. (2004), The nature of lead users and measurement of leading edge status, *Research Policy*, 33, 2: 351-362.
- Muniz A.M. Jr., O'Guinn T.C. (2001), Brand Community, *Journal of Consumer Research*, 27: 412-432.
- Nambisan S. (2002), Designing virtual customer environments for new product development: toward a theory, *Academy of Management Review*, 27, 3: 392-413.
- Neuendorf K.A. (2002), *The Content Analysis Guidebook*, Beverly Hills, California, Sage Publications.
- Piller F.T., Walcher D. (2006), Toolkits for idea competition: A novel method to integrate users in new product development, *R&D Management*, 36, 3, pp. 307-318.
- Prandelli E., Verona G. (2006), *Collaborative Innovation*, Carocci, Roma.
- Prandelli E., Verona G., Raccagni D. (2006), Diffusion of web-based product innovation, *California Management Review*, 48, 4: 109-135.
- Sawhney M., Prandelli E. (2000), Communities of creation: managing distributed innovation in turbulent markets, *California Management Review*, 42, 4: 24-54.
- Sawhney M., Prandelli E., Verona G. (2005), Collaborating to create: The internet as a platform for customer engagement in product innovation, *Journal of Interactive Marketing*, 19, 4.
- Schouten J.V., McAlexander J.H. (1995), Subcultures of consumption: an ethnography of the new bikers, *Journal of Consumer Research*, 22: 43-61.
- Sebastiani F. (1999), *Machine Learning in Automated Text Categorization*, ACM Computing Surveys.
- Sitz L. (2008), Beyond semiotics and hermeneutics. Discourse analysis as a way to interpret consumers' discourses and experiences, *Qualitative Market Research: An International Journal*, 11, 2: 77-91.
- Thompson C. (1997), Interpreting consumers: a hermeneutical framework for deriving marketing insights from the text of consumers' consumption stories, *Journal of Marketing Research*, 34, 4: 438-456.
- Urban G., Von Hippel E. (1988), Lead user analyses for the development of new industrial products, *Management Science*, 34, 5: 569-582.
- Verona G. (2000), *Innovazione continua: risorse e competenze per sostenere il vantaggio competitivo*, Egea, Milano.
- Von Hippel E. (1986), Lead users: a source of novel product concepts, *Management Science* 32, 7: 791-805.
- Von Hippel E. (2005), *Democratizing Innovation*, The MIT Press, Cambridge, MA.
- Winter S. (1987), Knowledge and competence as strategic assets, in Teece D. (ed.), *The competitive challenge – Strategies for industrial innovation and renewal*, Ballinger, Cambridge.

- Zaltman G. (1997), Rethinking market research: putting people back in, *Journal of Marketing Research*, 34: 424-37.
- Zanasi A. (2005). *Text mining and its applications to intelligence, CRM and Knowledge management*, WITpress.