Paper to be presented at

DRUID15, Rome, June 15-17, 2015

(Coorganized with LUISS)

The Dynamics of Category Labels during Industry Emergence: Evidence from ?Smartphones?

Diego Zunino
Boston University
Dept. of Strategy and Innovation
dzunino@bu.edu

Stine Grodal
Boston University
Strategy and Innovation
grodal@bu.edu

Fernando F Suarez
Boston University
Strategy & Innovation
fsuarez@bu.edu

Abstract

Recently the role of categories in shaping market dynamics has become an important area of research. Part of this research stream has focused on the role of category labels?the first and most visible instantiation of categories. To date, little is known about how category labels evolve and why some are selected over others. We contribute to the category literature by integrating insights from linguistics, to understand the formation and selection of category labels. We investigate the generation and use of category labels over time by producers in the smartphone industry between 2000 and 2010. We identify empirical regularities in the evolution of the structure and content of category labels: some characteristics show greater turbulence, while others are stable over time. We use econometric matching techniques to show that spatially unified compounds?that is compounds which omit the space between the two compounded elements (e.g. ?smartphone? versus ?smart phone?)? gained better traction in the early phase of the industry.

Jelcodes:M00,-

The Dynamics of Category Labels during Industry Emergence: Evidence from “Smartphones”

Abstract

Recently the role of categories in shaping market dynamics has become an important area of research. Part of this research stream has focused on the role of category labels—the first and most visible instantiation of categories. To date, little is known about how category labels evolve and why some are selected over others. We contribute to the category literature by integrating insights from linguistics, to understand the formation and selection of category labels. We investigate the generation and use of category labels over time by producers in the smartphone industry between 2000 and 2010. We identify empirical regularities in the evolution of the structure and content of category labels: some characteristics show greater turbulence, while others are stable over time. We use econometric matching techniques to show that spatially unified compounds—that is compounds which omit the space between the two compounded elements (e.g. “smartphone” versus “smart phone”)— gained better traction in the early phase of the industry.

Introduction

Recent management research has started focusing on the role of categories in markets. Categories are socially constructed partitions that are used to group objects that are perceived to be similar (Bowker & Star, 2000). The first stream of work on categories studied in impact of category affiliation on firm performance (Zuckerman 1999; Kennedy, 2008; Pontikes, 2012). Other studies have examined the role of stakeholders' political power in the adoption or persistence of categories (Lounsbury & Rao, 2004), while others have noted that categorical evolution is far from linear (Navis and Glynn, 2010; Kennedy et al., 2010). Recent work has begun to investigate the co-evolution of categories and technological designs throughout the industry life cycle (Grodal et al., 2014). Yet, there are still many unanswered questions pertaining to category dynamics during the industry life cycle.

Kennedy and Fiss (2013) acknowledge that progress has been made in understanding the role of categories in cross-sectional studies, but they call for more empirical research focusing on categorical dynamics , particularly to understand why certain category labels persist while others fall out of favor. Such analyses might extend the insights coming from case studies (Nigam & Ocasio, 2010) and history analysis (e.g., Kuilman & Wezel, 2013), which have shown that the meaning of categories change over time. In particular, existing studies have tended to focus on a few category labels and tracked their evolution over time (Khaire & Wadhwani, 2010; Navis & Glynn, 2010). Instead, we propose to study the full set of category labels that were created within an emerging industry to provide insight into the kinds of labels that dominate at different points in time during the industry life cycle.

To conduct our research, we use category labels as our unit of analysis. Category labels are the first instantiation of categories: they are words or phonemes used to describe a category (Grodal et al., 2014). Working as boundary objects, they can cross organizational and cultural boundaries and thereby transmit their meaning across time and space (Granqvist et al., 2013). As visible manifestation of categories, category labels evoke expectations in audience members, assisting them to navigate the categorical space which, particularly in emerging market spaces, can be confusing (Kovacs & Johnson, 2013). New category labels are created when firms and other stakeholders invent new ways of talking about their products or their particular view of the industry.

We currently do not understand how labels are constructed, how they change over the industry life cycle, and what characteristics make them more appealing to stakeholders. So far, few empirical studies have focused on the category label as unit of analysis and even fewer have tracked the evolution of labels over time. Our study addresses this deficiency. We also contribute to developing a theory of category evolution by integrating insights from linguistics into current understanding of category dynamics within the management literature. This former field—in particular psycholinguistics—has devoted particular attention to the classification (Booij, 2005; Guevara & Scalise, 2009) and analysis (Juhasz et al., 2003; Liu & McBride-Chang, 2010) of compounding. A compound is "the simple concatenation of any two or more nouns [or other words] functioning as a third nominal" (Downing, 1977: 810). We particularly borrow from Juahsz et al. (2003) notion of spatially unified compounding. We argue that there is an association between the traction that labels receive within the market space and the labels' structure; more specifically, labels that use spatially unified compounding will tend to be associated with higher traction in the markets.

The contribution that this paper makes to our understanding of categories is twofold. First, by focusing on a large market space that witnessed the creation of many category labels—the smartphone industry—we provide a detailed account of how category labels are both created and how the labels' structure and content develop over time. Second, we bring to the existing literature of categories the concept of spatially unified compounding as an important "signaling" mechanism that category sponsors may use to promote their category label. We argue that labels that use spatially unified compounds signal that the category is more mature and its boundaries are firmer. Using regression analysis and econometric matching techniques, we show that spatially unified compounds indeed have greater traction than compounds that are spatially separated.

The paper is organized as follows: section 2 reviews the relevant literature, section 3 describes the trends in the smartphone industry, section 4 presents an econometric evidence about the relationship between spatially unified compounding and labels' market traction, while section 5 discusses and concludes.

Theoretical Background

Categories and Market Dynamics

Recently management scholars have begun to investigate how categories shape market dynamics (Zuckerman, 1999). Categories both help market participants classify social objects (Zuckerman, 1999) and reduce ambiguity when considering similar products (Navis & Glynn, 2010; Pontikes, 2012; Durand & Paolella, 2012). Categories are arguably the most prominent sociocognitive concept used to study the evolution of emerging spaces; however, other concepts have been used in the literature, such as cognitive taxonomies (Porac & Thomas, 1990), field frames (Lounsbury et al., 2003), schemas (Bingham & Khal, 2013), or technological frames (Kaplan & Tripsas, 2008). Categories are socially constructed by multiple stakeholders within the market. Companies might create and sponsor categories, but critics and consumers can influence which categories eventually become accepted. For instance, activism and media can exert social pressure that is related to the rise (Lounsbury et al., 2003) and fall (Piazza & Perretti, 2015) of categories.

Category labels are signs or symbols that transmit meaning over time and place (Granqvist et al., 2013). The more stable the pattern of features associated with a category label, the more likely the audience will better recognize and understand the respective category (Weick, 1995; Rosa et al., 1999; Bowker & Star, 2000; Kennedy et al., 2010). Category labels evoke expectations in the audience, and they assist them when navigating the organizational space (Kovacs & Johnson, 2013). When category labels are ambiguous, they can cause confusion, which can hurt the performance of organizations that affiliate with those categories (Pontikes, 2012). However, the process through which category labels get infused with meaning is under-researched (Grodal et al., 2014). A better understanding of how category labels evolve can provide new insights about how this process unfolds.

Few recent studies theorize about the evolution of category labels during the industry life cycle (Grodal et al., 2015; Suarez et al., 2015). These studies suggest that when a new industry emerges, different stakeholders introduce an increasing number of category labels to describe the new technology. This creates contestation and competition among category labels. The contestation and growth in the number of category labels does not cease until the emergence of a dominant category label—"the conceptual schema that most stakeholders adhere to when referring to products that address similar needs and compete for the same market space" (Suarez et al., 2015: 437). After this point in time, the number of category labels begins to shrink, the industry gradually converges around the dominant label, and that category label increasingly defines the category with greater depth and more unambiguous boundaries. Therefore, while extant research has investigated how many labels evolve along time and the industry lifecycle, little is known about which particular labels evolve, and why.

This paper contributes to filling this gap by studying how the structure and the content of category labels evolve over time. In studying the content and the structure of category labels, we integrate the literatures of categories, industry lifecycle, with that of cognitive linguistics (in particular the work on compounding). As we argue below, spatially unified compounds (Juahsz et al., 2003) act as a "signaling" mechanism that is associated with greater traction of a category label.

The empirical analyses of categorical dynamics have been so far limited. Often, empirical research has kept the category constant over time and studied the behavior of its actors (Navis & Glynn, 2010). While case studies (e.g., Nigam & Ocasio, 2010) and history event analysis (Kuilman & Wezel, 2013) have provided some information on how category labels evolve, we still miss the key dimensions of diffusion and degree of adoption.

In this paper, we propose a hybrid methodology. To capture the nuances of the dynamic change of category labels, we will graph important aspects about the structure and the content of the category labels over time. To investigate the relationship between spatially unified compounds and the traction of category labels, we will use econometric techniques for dealing with count data, mitigating potential endogeneity issues by applying panel data and non-parametric matching techniques (Iacus et al., 2011).

Compounding in linguistics

Despite the increasing interest in categorization, when dealing with categorical dynamics, few management scholars have drawn on insights gained in linguistics. Yet, categories and categorical labels are inherently linguistic in nature. According to linguistic taxonomy (Booij, 2005), every category label with more than one word can be classified as a compound, as in "personal computer", "typewriter", or "airplane". Compounding is a linguistic technique to create new words by recombining existing words. Compounding is the most common way for industry stakeholders to create new category labels to refer to their products (Grodal et al., 2014). In addition to compounding, a less common but still used linguistic technique to create new words and labels is derivation, or the transformation of an existing word by modifying its tense or ending, such as in "browser" which is derived from the verb "to browse" (Guevara & Scalise, 2009).

Compounding involves the juxtaposition of two or more lexemes in order to define a new object. Compounds are formed by a "head" and one or more modifiers. A head is a word, most generally a noun, that defines the nature of a phrase-- such as “computer” in the “personal computer” label. Modifiers add meaning to the compound, such as “personal” in “personal computer.” Compounds can be classified by their morphology. How to classify compounds is a sub-field in linguistics (Bisetto & Scalise, 2005). Obviously, an exhaustive review of this literature goes beyond the scope of the paper, but for our purposes here it is important to retain the key difference between a head and a modifier in compounds.

A head noun, in turn, can give rise to what linguists call endocentric or exocentric compounds. When a compound has a clear head in which one of the elements of the compound carries most of the semantic meaning they are called endocentric. For example in the compound “smartphone” it is “phone” that carries most of the semantic meaning. In contrast if both elements of the compound carry equal weight they are called exocentric (Bloomfield, 1933). For example, “screwdriver” is an exocentric compound, because a screwdriver is not a kind of “driver” nor a kind of “screw”.

When we look at the modifier we can distinguish between attributive/appositional (Scalise et al., 2005) and relational (Gagné & Spalding, 2008) compounds. Attributive and appositional compounds are those whose modifers adds meaning in the form adjective-noun (e.g., “smart device”) or noun-noun (e.g., “computer phone”) to the head noun. For example, “smart device” is an attributive compound, because its modifier defines one characteristic of the head noun. In contrast in a relational compound, the modifier explains an elicited relationship between itself and the head noun. For example, “lap top” compound suggests that the object stays on the laps.

The literature on compounds within linguistics pays little attention to the evolution of compounds over time. The only dynamic reference is related to spatially unified compounding (Juashz et al., 2003), a feature of compounds where the lexemes are either merged or joined by a hyphen. According to the extant literature in linguistics, there is no clear pattern in the evolution of compounds from distinct words to spatially unified compounds (Juashz et al., 2003: p. 224).

By using a spatially unified compound, producers signal that a category is more solidified and its boundaries are stronger. In contrast, if they still include a space in the category label then they signal that the adjective is still moderating the noun. The unifying technique can be associated to the deepening of categories as the industry matures, as hypothesized by Grodal et al. (2014). For example, using the compound "smartphone" signals that the category has stronger boundaries than using the compound "smart phone". Therefore, we hypothesize the following:

> Hypothesis 1. Labels that use spatially unified compounds are associated with higher market traction than labels that use a different linguistic structure.

The Evolution of Category Labels

Setting: Smartphone Industry

We choose to study category labels within the smartphone industry. The common trait of smartphones is the presence of an advanced operative system, the possibility to browse the internet, run applications, and make phone calls on the same device. In 1992 IBM introduced the first device containing these three elements. The first European device with these elements was the Nokia 9000, which hit the market in 1996. In the beginning of the industry these devices were, however, not referred to as "smartphones". In fact, Ericsson introduced the term "smartphone" in 1997 when it launched its Ericsson GS88 model. Operating systems play a crucial role in the smartphone industry. Early on Nokia, Psion, Motorola, and Ericsson jointly developed a smartphone operating system. The entry of Blackberry and Apple into the market disrupted this existing market order. In particular Blackberry and Apple exclusively sold smartphones and had no feature-phone offerings like their other competitors. By 2013, sales of smartphones overtook the sales of traditional feature-phones (Cecere et al., 2014), The smartphone industry represents an ideal setting for addressing our research question for three reasons. First, it is relatively young, which facilitate accurate data collection. The first models appeared in 1996, and for the period between 2000 and 2010 we have been able to retrieve all the category labels from almost every press release. Second, there are multiple heterogeneous producers. There are de novo entrants, and entrants from closer and more distant technological products –e.g., phone industry and PC industry, respectively. This plurality of participants is likely to generate a plethora of different category labels. Furthermore, because there are multiple producers we are likely to observe variance in category creation and affiliation even within each type of producer. Third, the fast pace of technological change creates uncertainty, which again might stimulate variance in the creation and affiliation with category labels. Within the smartphone industry the meaning of category labels are likely to be overlapping or ambiguous. These three aspects makes the smartphone industry appealing for the study of the evolution of category labels over time.

Data Collection

We collected category labels from producers' and operators' press releases. When producers create a new technology they most often create a press release in order to communicate and position the new product to consumers and other market stakeholders (Maat, 2007). Moreover, existing research on categories have used press releases as a data source (Kennedy, 2008; Navis & Glynn, 2010; Pontikes, 2012). We collected information from 382 press releases introducing a new smartphone in the North American and European market between 2000 and 2010. These press releases come mostly from the 31 different producers. If we were not able to associate a new product release with a producers' press release we relied on the press release from the operator instead.

From each press release, we extract the category labels that the producer or operator use to describe the new technology. We provide here an example from the first paragraph of the 2008 press release introducing Nokia N97[1].

> "Barcelona, Spain -- Nokia today unveiled the Nokia N97, the world's most advanced mobile computer, which will transform the way people connect to the Internet and to each other. Designed for the needs of Internet-savvy consumers, the Nokia N97 combines a large 3.5" touch display with a full QWERTY keyboard, providing an 'always open' window to favorite social networking sites and Internet destinations. Nokia's flagship Nseries device introduces leading technology - including multiple sensors, memory, processing power and connection speeds - for people to create a personal Internet and share their 'social location."

From this paragraph, we can isolate two category labels. The first one is "mobile computer", and the second is "Nseries device". We proceeded in a similar fashion for 392 press releases over 11 years.

Category labels are words that "transmit meanings across time and space" (Granqvist et al., 2013): at the beginning they are not meaningful by themselves, over time they get infused with meaning and they are associated with a category. Over time stakeholders within an industry begin to prefer some categories over others, whereby some categories begin to use used more frequently whereas others fall out of favor. The category label that wins the battle becomes the "dominant category label" (Suarez et al., 2014, Suarez & Grodal, 2014).

Stakeholders, not only producers, use category labels when referring to the new technology. There are cases where these stakeholders can generate new category labels as well. For example, the category label "impressionism", which is commonly used to refer to a particular genre of 19th century paintings, comes from a satirical review by Louis Leroy, which appeared in the Parisian newspaper "Le Charivari" on April 25th, 1874. Despite this and other notable exceptions, manufacturers create the lion's

[1] http://company.nokia.com/en/news/press-releases/2008/12/02/desktop-laptop-pocket-the-era-of-the-personal-internet-dawns-with-the-nokia-n97

share of category labels within an emerging industry, because producers tend to be more influential, and have the strongest incentives to create new category labels. For this reason, we limit our attention to producers and operators in this paper.

We found that producers often use more than one category label within a given press release. Figure 1 shows the evolution over time: we not only report the averages, but we also plot a smoothing trendline using a third-order polynomial approximation. Since the beginning of the industry, companies use on average more than three category labels per press release The range goes from more than three (3.5) to almost five (4.91). Figure 1 suggests that the development in the number of category labels that producers use follows an inverted U-shape. That is, producers use few category labels per press release during the beginning of the industry but, over time, they use more category labels in their press releases; the number of average labels used reaches a peak, and declines again. This pattern seems in line with theories in the category literature, where scholars propose an inverted U-shape relationship between the total number of category labels that are used within an industry and time (Suarez et al., 2014). Firms increase the number of category labels per press release over time, suggesting that they are experimenting with how to refer to the novel industry. The peak in the number of the labels per press release is 5; after that, this figure decreases to 3.5, suggesting that over time experimentation decreases and firms begin to converge on a more limited amount of labels to use to refer to the novel industry.

Insert Figure 1 here

In the next sub-sections we describe the empirical regularities we found in our unique dataset. We define two levels of analysis: The structure level and the content level of a label. In the first level of analysis, we observe the characteristics of the label in terms of linguistic structure (e.g., spatially unified compounding). In the content level of analysis we focus our attention on the modifiers of the labels and the particular meaning of each category label.

Structure of the Labels

We first analyzed the particular structure of a category label. We coded each single word in a label according to their type, adjective (A) or noun (N). We used as baseline the traditional parts of speech (verbs, adverbs, and pronouns), we omitted interjections, we joined prepositions with conjunctions, and we added two categories for acronyms (coded as 1) and numbers (coded as 0). For example, we coded the label "smartphone" as "A-N", the label "pocket device" will be coded as "N-N", and the label "Android-enabled device" will be "N-A-N". Eventually, we coded 128 different combinations of structures. Figure 2 shows the five most used structures. We observe that the majority of category labels are a recombination of two types, nouns and adjectives. No other parts of speech appear in the figure. From the graph, we can observe that producers prefer simpler labels: category labels with just a noun, or with a noun and an adjective account between 35% and 50% of all the category labels. Notwithstanding, between 2003 and 2005, there seems to be more contestation in the type of structure preferred, with alternative labels taking over (e.g. the structures "noun-adjective-noun" and "noun-noun"). Eventually, Figure 2 suggests that the structure contestation seems to resolve in favor of two main structures: "adjective-noun" and simply "noun". The prevalence of adjective-noun compounds is consistent with existing research which suggests that 2-word compound category labels can better manage the tension between familiarity and originality that characterizes many successful labels (Grodal et al., 2015).

Insert Figure 2 here

Figure 3 plots the average number of words for each label.. Until 2002, category labels are composed by an increasing number of words – reaching a peak of almost 3 (2.8) words per label on average. After that year, the number of words decreases to around 2 (2.2) words per label. This result suggests that in the initial phase of categorical evolution, the reigning uncertainty prompts producers to use more words in an attempt to add meaning to their labels. However, over time labels get infused with meaning and the market increasingly prefer 2-word compounds, which as noted above have been proposed as an efficient way to convey both familiarity and originality.

Figure 4 shows the yearly share of a particular type of labels whose structure is one of “spatially unified compounds”. From linguistics perspective, all the category labels in the dataset can be treated as compounds, except those labels with only one word. Linguistic scholars have long debated how to classify the different types of compounds (see for reference Bisetto & Scalise, 2005). As noted earlier. one notable typology divides compounds between endocentric and exocentric (Booij, 2005).

Our data contains almost exclusively endocentric compounds. For example, the label “smart device” is endocentric, as we can clearly understand from the head noun that the labels directly refers to a device. A category label such as “handheld” is exocentric, as we cannot infer from the head noun “hand” that the label is referring to a phone. The label “handheld” is also the most notable exocentric compound in our sample. It became popular in 2002, and started to decline until 2006.

We focus here on spatially unified compounds. Their study is a relevant topic for linguists: Liu & McBride-Chang (2010) call for studies to understand whether earlier knowledge of spatially unified, or explicit, compounding structure results in more widespread usage. Despite of the fact that, in the English language, most compounds do not follow a particular structure, there is a potential evolutionary path. As the co-occurrence of two words increases, the two words tend to appear either joined by a hyphen or joined into a single word – these later ones are referred to as spatially unified compounds (Juhasz et al.,2003). We believe that category labels with a spatially unified compounds structure can be a form of categorical “signaling” – intentional or unintentional – by a firm. Spatially unified compounds may signal that the type of product described belongs to a more mature category with firmer, more solidified boundaries.

We coded spatially unified compounds accordingly. Figure 4 shows clearly that their use increases over time. Their use peak up to more than 60% in 2002, then drops to 30% in 2008, and keeps growing afterwards reaching 50% in 2010. The signaling nature of spatially unified compounds can explain the peak in the earliest stage. It means a multitude of attempts to let stakeholders make sense of a new category. This empirical evidence is in line with the stage of linguistic recombination as in Grodal et al. (2014). The takeoff in spatially unified compounds starting in 2008 is most likely related to the emergence of a dominant category label in the industry: “smartphone”.

Insert Figure 3 here

Insert Figure 4 here

We next analyze the evolution of “head nouns”. In linguistics, a “head” is the word that defines the nature of the phrase, while the “modifier” is the rest of words that add meaning to the phrase. For example, in the compound “handset” the head noun is “set”, while in “camera phone” the head noun is “phone”. As noted earlier, linguistic authors (e.g. Booij, 2005) sustain that a modifier can provide attributive/appositive or relational meaning. In our dataset, most of the labels are of attributive type, with a limited number of relational compounds, some notable examples among them are “handheld” and “handset”. Relational compounds’ diffusion does not go much beyond 10% of the share of labels used in a year.

We investigate the head nouns because they illustrate the different stems from which manufacturers choose their labels. In essence, the head noun that ends up as part of the dominant category label identifies the perceived nature of the product. In the smartphone industry, manufacturers used 37 head nouns for 390 category labels. In Figure 5, we report the first 8 head nouns that were used most frequently. Out of the 37 different head nouns, these 8 most-used ones account for 93% of all citations. The figure shows that there was contestation between the different head nouns up to 2003, after then, contestation only remains between “phone” and “device”. The former is a head noun evoking the prior technology from which the new industry sprung, while the latter is a general word with broad meaning.

The diffusion and prevalence of the head noun “phone” can simply be due to path dependency. Most of the early smartphone producers used to produce the previous generation of more basic, (feature) phones. Despite the substantial technological distance, it is likely that they perceived the products based on the new technology in categorical proximity to the prior-technology products. By referring to the technology from which the new industry spins off, the head noun “phone” offers producers a bridge of familiarity that can ease the transition for customers and other stakeholders.

The persistence of “device” as head noun can simply be due to producers using this general word in press releases in order not to repeat a more important category label that they may be promoting. Also, the generality and the neutrality of the head noun “device” makes it easier to combine with a larger number of modifiers. The figure suggests that some manufacturers may have used the noun device as a form of “wait and see” until they could elucidate which label would emerge as the dominant category in the industry.

Insert Figure 5 here

To wrap up, we note the contestation in the structure of category labels we observe is consistent with recent research that proposes that labels are created in emerging industries through “recombination” of existing linguistic elements; we indeed observe this in the early stage of our industry, peaking in 2002 and 2003. As argued by Suarez et al. (2014), this contestation among category labels seems to precede the contestation of technological designs and the emergence of a dominant design in the industry which many authors consider to be Apple’s iPhone, first released in 2007.

Content of the Labels

In this subsection, we analyze the particular meaning labels get infused with. We do this by analyzing the type of modifiers used in the compounds, and study how modifiers evolve over time. We distinguish three broad types: modifiers that reference a technology, those that reference a trademark, and those that reference a particular segment. We further divide modifiers that reference a technology into two sub-categories: those that reference a technological generation (such as 2G or 3G wireless), and those with reference an operative system (such as Android or Windows).

Figure 6 represents the trend in the share of labels that reference a general technology. By referencing a general technology we mean those category labels whose modifier refers to technological features, either hardware or software, such as "Bluetooth-enabled device", "camera phone", or "email device". This type of category labels is very common: the share of category labels that reference a technology varies from 10% to 45% over time. In general the trend is quite stable, besides the peaks of 2002, 2004, and 2008. The cyclical trend roughly matches the diffusion of particular technologies over time: for instance, the 2002 peak corresponds primarily to CDMA; that of 2004 to 3G and 2G-edge; and 2008 to the entry of Android and the renewed effort of Microsoft to promote Windows mobile.

Insert Figure 6 here

Figure 7 shows the category labels that reference a technological generation. These are labels that reference a particular version, for example "3G device." Whenever a producer chooses to use a category label that refers to a technological generation, there is a tradeoff between short-term traction while the technology is perceived as novel and longer-term obsolescence if the technology begins to be perceived as "dated". For example, this happened to 3G labels with the advent of 4G. When 4G emerged, some producers reacted by trying hybrid levels not exempt from ambiguity, such as "3G/4G phone."

In general, these technology-related category labels are not used very often in our industry. Their use peaks at under 10%.

Insert Figure 7 here

Figure 8 shows the situation of category labels that reference an operative system, such as Windows Mobile, Symbian, or Android. Examples of these category labels are “Windows mobile smartphone”, “WebOS device”, or “4G Android-powered device”. Leveraging a particular operative system can be useful to compensate for the ambiguity about the product. We observe a consistent correlation between category labels that mention an operative system and those that use suffixes such as “-powered” and “-enabled“. In figure 9 we reported the shares of category labels that use the above mentioned suffixes: the trend is the same as that observed for the case of labels that reference an operating system.

In the early stage, leveraging the fact that a device uses well-known operating system, such as Windows, can increase acceptance of a new product. In fact, the consumer may be already exposed to Windows in a different industry like PCs. In a later stage of the industry lifecycle, e.g., after 2006 in our industry, leveraging an operative system can be a way to differentiate the product when a dominant category label or a dominant design emerges. In our data, the use of labels referring to an operative system follows a U-shape pattern. It is interesting to note that the use of these labels reaches a bottom around 2004, when there is a peak in contestation about technological features and technological generations, which later gives way to a widespread diffusion of prominent operative systems, chief among them Android..

Insert Figure 8 and 9 here

In Figure 10 we move to the analysis of category labels that contain a trademark. Trademarks can belong to the producer that emits the press release (e.g., “Blackberry device”, “T-mobile phone”), or it can refer to some trademarked feature or technology (e.g., “Windows phone”, “Bluetooth device”). Using a trademark in a categorical label can be risky. It can limit the traction of the label because by law trademarked words cannot be freely used by other stakeholders. Moreover, if the trademark widely diffused it can have diluting effects on the producer’s brand. In our data, the diffusion of trademark category labels is quite stable over time; there seems not to be any particular pattern for this type of labels over the industry lifecycle.

Insert Figure 10 here

In figure 11 we present the last notable characteristic in modifiers: the reference to a segment. When the categorical space begins to crowd up, a possible strategy can be to attempt to carve out a new niche through the use of segment-specific category labels. For example, there can be smartphone models for fashion-sensitive people (as suggested by many co-branding experiences, such as Motorola–Dolce & Gabbana, Samsung-Armani, or LG-Prada), or there can be smartphones that clearly target the business segment (such as the first Nokia Communicators, and the early Blackberry models). Some category labels explicitly target specific category of users, as in “fashion phone”, “business-optimized device”, or “mobile office.”

While market segmentation is very common in business, it does not seem that common as a labeling strategy. The share of segment-oriented labels is very limited, peaking in 2005 at 5.5% of the total. The peak happens immediately after 2004, right after the contestation for structure and technological features. Eventually, segment labels are only one of the many ways to infuse meaning to categories, and their use fades abruptly as soon as the dominant category label emerges.

Insert Figure 11 here

The traction-premium of spatially unified compound category labels

The goal of this section is to find initial evidence of how the structure of category labels is related to their adoption in an industry. Ultimately, our findings may shed light on how different label strategies at the firm level could influence the categorical evolution in the industry and the performance of the firms that sponsor the different type of labels. A well-informed category label strategy can potentially help firms have their labels more diffused, reduce ambiguity around their products, and therefore, to improve overall firm performance.

We study here the association between a label's traction and the use of spatially unified two word compounds in its structure. We claim that spatially unified compounding can augment the traction of a category label and make it more likely to become the dominant category in an industry. Our unit of analysis is category label-year. We operationalize traction as following: we use our dataset of press releases, and we count for each year how many press releases cite a particular label. We define spatially unified two word-compounds as those compounds that have a hyphen or no space between two words. We operationalize it as a dummy variable that takes the value of one if the label is a spatially unified two-word compound. We add control variables for label age, the other label characteristics described and analyzed in the earlier section, and year dummies. The regression equation is the following:

$$\log(cites_{it}) = \alpha + \beta SpUComp_i + \Gamma X_{it} + \tau + \kappa + u_{it}$$

Because we have to deal with count data, we use a log-linear regression model, and we use a logarithmic transformation of the traction variables. The coefficient β represents the percentage difference in the counts of the label when its structure corresponds to a spatially unified two-word compound. X is a vector of label-specific characteristics of the category label, τ is a year dummy, and κ is a head noun dummy, and u_{it} is an unobserved component.

Because the choice to use a spatially unified compound is non-random, the regression is likely to suffer from endogeneity. Omitted variable bias is the major concern, as there may be some variables that we were not able to measure and that affect label traction, such as the auditory pleasure in the sound the words. We are not able to fully correct for possible endogeneity issues, but we try to mitigate these in two ways: first, we use random effects panel techniques; second, we perform matching algorithms.

We use random effects as panel technique because our explanatory variable is label-specific and therefore it could not be estimated using fixed effects. Therefore, we have to rely on the assumption that label specific effects are uncorrelated with the other covariates. Given our data, this assumption can be strong. Therefore, we proceed with our second estimation technique, matching. We use coarsened exact matching, a nonparametric technique (Iacus et al., 2011) that allows for matching based on blocks rather than exact matching. Matching estimators rely on the conditional independent assumption, that is, that the choice of spatially unified compounding for a label is almost random given the observables. Formally:

$$\left(\log(cites)_{SpU}, \log(cites)_{NoSpU}\right) \perp SpUComp|X$$

As Table 1 shows, when we separate the labels with spatially unified compounds in the unmatched sample, they seem to have systematically different characteristics from the rest of the labels. Spatially unified labels tend to be shorter, and refer less to technologies and technology generations, while referring more to lifestyle and containing more derivations. After matching, the unbalance almost disappears, with the only exception of “derivation”.

Table 2 shows the results of the pooled cross-section, random-effects, and matching regressions. Model 1 is an OLS model with full sets of dummies for head nouns and years; model 2 mitigates endogeneity through the use of random effects; model 3 is an OLS regression on the matched sample; and model 4 is an OLS regression that uses an alternative matching algorithm that includes also matching based on the head nouns. We first analyze the controls and we will then turn to the explanatory variable.

Age seems to be associated with traction in a U-shape relationship, and this finding seems quite reasonable: category labels may have a harder time to get traction initially but their adoption improves as they get infused with meaning over time. Category labels citing a technological generation seem positively associated with cites but not at conventional significance levels in the first two models. When the sample is more balanced through matching, their coefficient become significant and these labels account for 21% to 61% more citations than analogous category labels, in models 3 and 4, respectively. The negative and significant coefficient of the technology reference dummy mitigates the previous positive relationship, as the two variables often co-occur. Category labels that refer to a technology get a discount from 37% to 50% in adoption in the matched samples. The reference to an operative system is positively associated to traction and it ranges in the matched sample from 47% to 58%.

Trademark and endorsement suffixes do not seem to be correlated with traction at any statistical significant level. Segment related labels are negatively related with traction, and it is not possible to test this coefficient in the matched samples as the matching algorithm returns no matched pairs. This is due to the limited availability of this kind of labels: the negative correlation is justified by their limited diffusion and the cyclicality observed over time for lifestyle trends. Category labels that use derivation seem to be positively related with traction, even if the significance of the coefficients is not stable when using Random Effects or the alternative matching algorithm. This positive association confirms the derivation and compounding hypothesis found in Grodal et al. (2014).

Eventually, we turn to the analysis of our main predictor, spatially unified compounding. The coefficients for models 1 and 2 are positive, small, but not significant at the conventional levels. However, when we balance the sample between category labels using spatially unified two-word compounds, the coefficient becomes statistically significant and larger, suggesting a difference ranging from 31% to 42% according to different matching techniques. In other words, the category labels that use spatially unified compounds are more widely adopted than other types of labels by 31% to 42%.. As we argued earlier, there appears to be a premium in the adoption of spatially unified compounds that may be related to the fact that these reflect sharper and more stable category boundaries.

Discussion and Conclusion

Earlier research contributed to increase our understanding about the role that categories play in different industries (Zuckerman, 1999; Pontikes, 2012). One stream of research focused on how categories rise (Lounsbury et al., 2003; Nigam & Ocasio, 2010) and fall (Piazza & Perretti, 2015), and how they gain legitimacy (Navis & Glynn, 2010). These kinds of studies are characterized by the analysis of one single category at the time and they do not explore the underlying dynamics in category labels. Studies about the evolution of category labels and the industry life cycle (Suarez et al., 2014; Grodal et al., 2014) are still limited to theory building but have not yet provided detailed empirical analyses. Surprisingly, the study of categories in management literature has for the most part ignored the fact that the phenomenon of labels creation and adoption is intrinsically a linguistic one.

We borrow from linguistic literature to build our arguments and guide our empirical analysis in this paper, and focus on the concept of compounding (Bloomfield, 1933). We analyze some label characteristics in light of the linguistic and cognitive linguistic literatures, such as the structure of the compound (Bisetto & Scalise, 2005; Guevara & Scalise, 2009), the type of head noun (Booij, 2005), and the use of spatially unified compounds (Juhasz et al., 2005). While we do not claim to position our paper as a linguistics paper, we believe that much can be learned about the dynamics of the evolution of labels by using linguistic lenses. In addition, our econometric section represents a novel approach in a field where most of the empirical results come from experiments.

This article contributes to existing literature in a threefold fashion. First, it is one of the few empirical studies that use category labels as the unit of analysis. Very few studies do that, especially with a dynamic perspective. Second, our study adds granularity to our understanding of labels' evolution by looking at patterns in label structure and content. Third, we point to an eventual role of agency in label evolution, by suggesting how different category strategies could help a firm influence the evolution of labels in its industry and potentially improve its performance. We specifically suggest that the use of spatially unified compounds can result in a signal to stakeholders that a given label is more mature and stable.

We believe that our result can consequently have important implications for practitioners. We showed that there are particular patterns in the evolution of category labels. While companies may currently pay attention to their brands and branding efforts, most practitioners do not have a clear understanding of the process by which category labels are formed and adopted. We believe that a more informed understanding of these processes can help practitioners carve an advantage in the market by being able to anticipate and even influence the evolution and meaning of industry labels. . For instance, we have shown that the adoption of spatially unified compounds leads to higher adoption of the label sponsored by a particular firm.

Limitations and further research

This article suffers of two limitations that are worth noting: first, we focus on a single industry and therefore cannot tell if there are different industry patterns depending on the characteristics of an industry. Second, despite our attempts to mitigate it, we may have not been able to fully eliminate the endogeneity problems, reason why we have abstained from drawing any strong causal conclusion.

As for the external validity concern, we believe that the smartphone industry is an ideal industry for the study of these categorical issues, since it is characterized by heterogeneous and numerous players, these players come from different backgrounds, and the pace of technological change is fast. Given the state of the art for the study of compounding, mainly based on experimental evidence, we believe that studying actual label adoption by producers provides insights that go beyond those than can be gathered from experiments. Nevertheless, further research can overcome this limitation by focusing on other industries.

For the endongeneity concerns, we are aware that the assumptions behind these models can be considered strong – as any assumption would be in this research setting. One could potentially try to replicate the quantities estimated by using field experiments. A field experiment would have a more robust external validity than a lab experiment, and through random manipulation one would be able to identify the effect. We leave this for future research.

We also hope that this paper can motivate further work that draws from linguistics to study categories and category labels. We explored the relationships between label characteristics and adoption: in a way, each of our graphs represents an opportunity for a more granular research. Our econometric exercise on spatially unified compounding wanted to bootstrap this approach based on regression analysis vis-à-vis experiments. While we focused on the internal characteristics of labels, further research can look also at the characteristics of the adopters.

References

Booij, G. E. (2005). The Grammar of Words: An Introduction to Linguistic Morphology. Oxford University Press.

Bloomfield, L. (1933). Language. New York: Henry Holt

Bisetto, A., & Scalise, S. (2005). The classification of compounds. Lingue e linguaggio, 4(2), 319-0.

Bowker, G. C., & Star, S. L. (2000). Sorting things out: Classification and its consequences. MIT press.

Cecere, G., Corrocher, N., & Battaglia, R. D. (2014). Innovation and competition in the smartphone industry: Is there a dominant design?.Telecommunications Policy.

Downing, P. (1977). On the creation and use of English compound nouns. Language, 810-842.

Durand, R., & Paolella, L. (2013). Category stretching: Reorienting research on categories in strategy, entrepreneurship, and organization theory. Journal of Management Studies, 50(6), 1100-1123.

Gagné, C. L., & Spalding, T. L. (2009). Constituent integration during the processing of compound words: Does it involve the use of relational structures?. Journal of Memory and Language, 60(1), 20-35.

Granqvist, N., Grodal, S., & Woolley, J. L. (2013). Hedging your bets: Explaining executives' market labeling strategies in nanotechnology.Organization Science, 24(2), 395-413.

Grodal, S., Gotsopoulos, A., & Suarez, F. (2014). The Co-evolution of Technologies and Categories during Industry Emergence. Academy of Management Review, amr-2013.

Grodal, S., Suarez, F., & Zunino, D. (2015) *“Why Do Category Labels Stick? Industry Evolution and the* Battle for Catego*rical Dominance”* working paper, Boston University.

Guevara, E., & Scalise, S. (2009). Searching for universals in compounding. In Universals of language today (pp. 101-128). Springer Netherlands.

Iacus, S. M., King, G., & Porro, G. (2011). Causal inference without balance checking: Coarsened exact matching. Political analysis, mpr013.

Juhasz, B. J., Starr, M. S., Inhoff, A. W., & Placke, L. (2003). The effects of morphology on the processing of compound words: Evidence from naming, lexical decisions and eye fixations. British Journal of Psychology, 94(2), 223-244.

Kaplan, S., & Tripsas, M. (2008). Thinking about technology: Applying a cognitive lens to technical change. Research Policy, 37(5), 790-805.

Kennedy, M. T. (2008). Getting counted: Markets, media, and reality. American Sociological Review, 73(2), 270-295.

Kennedy, M. T., & Fiss, P. C. (2013). An ontological turn in categories research: From standards of legitimacy to evidence of actuality. Journal of Management Studies, 50(6), 1138-1154.

Kennedy, M. T., Lo, J., & Lounsbury, M. (2010). Category currency: A framework for analyzing the effects of meaning construction process. Research in the Sociology of Organizations, 31, 369-397.

Khaire, M., & Wadhwani, R. D. (2010). Changing landscapes: The construction of meaning and value in a new market category—Modern Indian art. Academy of Management Journal, 53(6), 1281-1304.

Kovács, B., & Johnson, R. (2013). Contrasting alternative explanations for the consequences of category spanning: A study of restaurant reviews and menus in San Francisco. Strategic Organization, 1476127013502465.

Kuilman, J. G., & Wezel, F. C. (2013). Taking off: Category contrast and organizational mortality in the UK airline industry, 1919–64. Strategic Organization, 11(1), 56-77.

Liu, P. D., & McBride-Chang, C. (2010). Morphological processing of Chinese compounds from a grammatical view. Applied Psycholinguistics, 31(04), 605-617.

Lounsbury, M., & Rao, H. (2004). Sources of durability and change in market classifications: A study of the reconstitution of product categories in the American mutual fund industry, 1944–1985. Social Forces, 82(3), 969-999.

Lounsbury, M., Ventresca, M., & Hirsch, P. M. (2003). Social movements, field frames and industry emergence: a cultural–political perspective on US recycling. Socio-Economic Review, 1(1), 71-104.

Maat, H. P. (2007). How promotional language in press releases is dealt with by journalists: Genre mixing or genre conflict?. Journal of Business Communication, 44(1), 59-95.

Navis, C., & Glynn, M. A. (2010). How new market categories emerge: Temporal dynamics of legitimacy, identity, and entrepreneurship in satellite radio, 1990–2005. Administrative Science Quarterly, 55(3), 439-471.

Nigam, A., & Ocasio, W. (2010). Event attention, environmental sensemaking, and change in institutional logics: An inductive analysis of the effects of public attention to Clinton's health care reform initiative. Organization Science, 21(4), 823-841.

Piazza, A., & Perretti, F. (2015). Categorical Stigma and Firm Disengagement: Nuclear Power Generation in the United States, 1970–2000. Organization Science.

Pontikes, E. G. (2012). Two sides of the same coin how ambiguous classification affects multiple audiences' evaluations. Administrative Science Quarterly, 57(1), 81-118.

Porac, J. F., & Thomas, H. (1990). Taxonomic mental models in competitor definition. Academy of Management Review, 15(2), 224-240.

Rosa, J. A., Porac, J. F., Runser-Spanjol, J., & Saxon, M. S. (1999). Sociocognitive dynamics in a product market. The Journal of Marketing, 64-77.

Scalise, S., Bisetto, A., & Guevara, E. (2005). Selection in compounding and derivation. AMSTERDAM STUDIES IN THE THEORY AND HISTORY OF LINGUISTIC SCIENCE SERIES 4, 264, 133.

Suarez, F. F., & Grodal, S. (2014). Hmm... What's That? How Categories Can Help Your Company Be a Tempo Mover in Nascent Industries. MIT Sloan Management Review, Forthcoming.

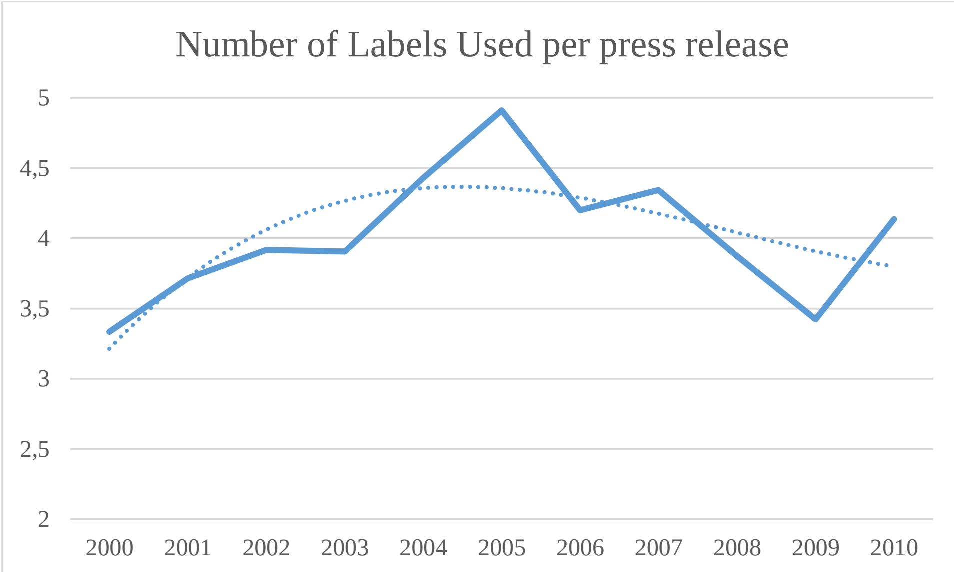
Suarez, F. F., Grodal, S., & Gotsopoulos, A. (2014). Perfect timing? Dominant category, dominant design, and the window of opportunity for firm entry.Strategic Management Journal.

Weick, K. E. (1995). Sensemaking in organizations (Vol. 3). Sage.

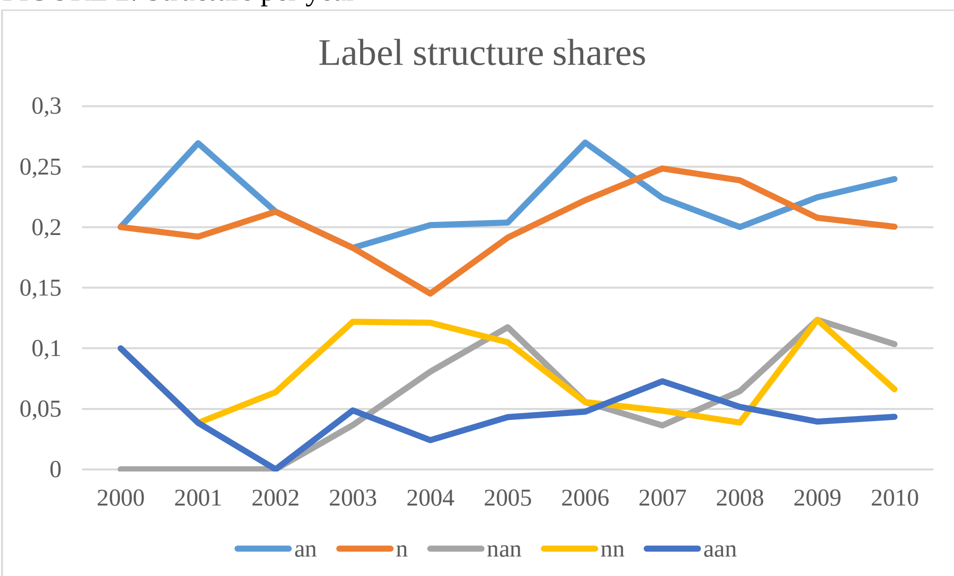
Zuckerman, E. W. (1999). The categorical imperative: Securities analysts anad the illegitimacy discount. American journal of sociology, 104(5), 1398-1438.

List of Figures and Tables

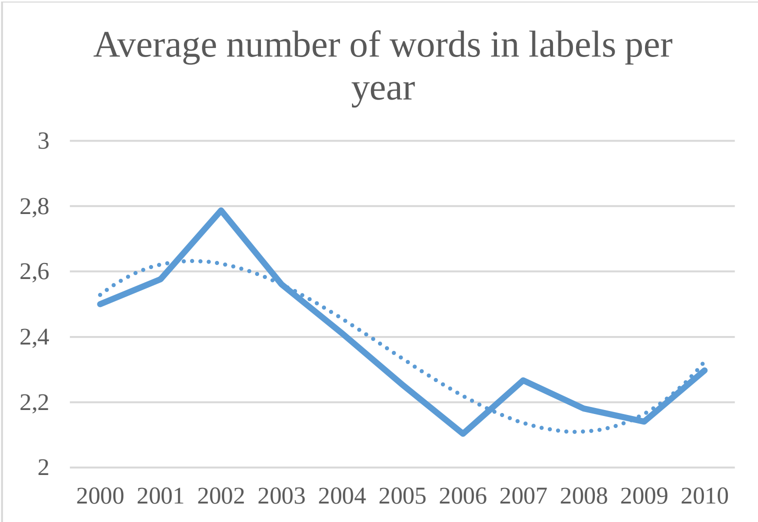
- FIGURE 1: Number of labels per press release



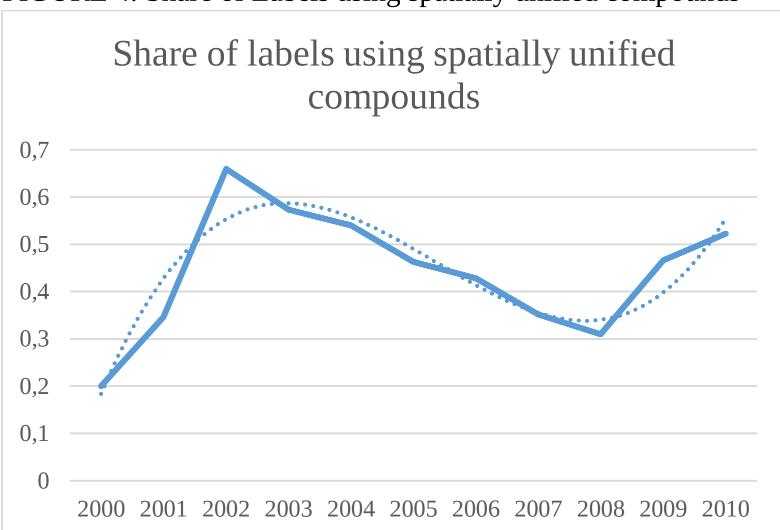
- FIGURE 2: Structure per year



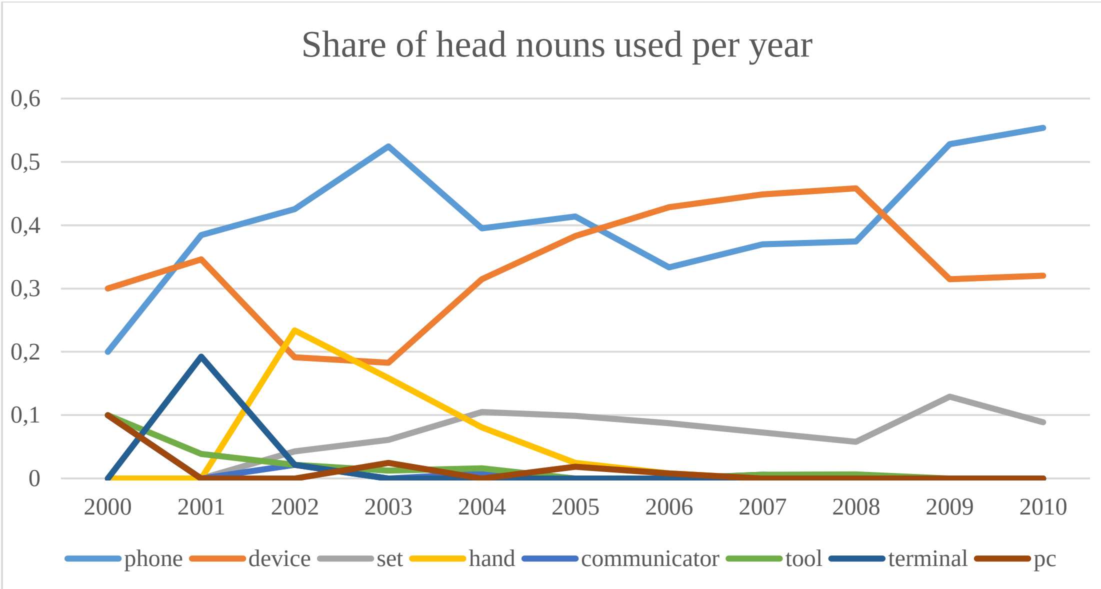
- FIGURE 3: Number of words per year



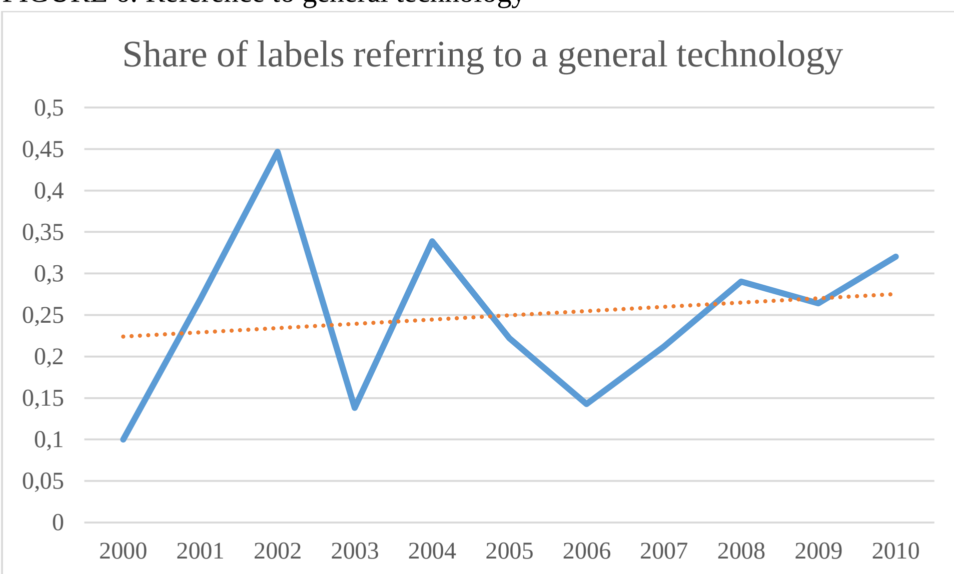
- FIGURE 4: Share of Labels using spatially unified compounds



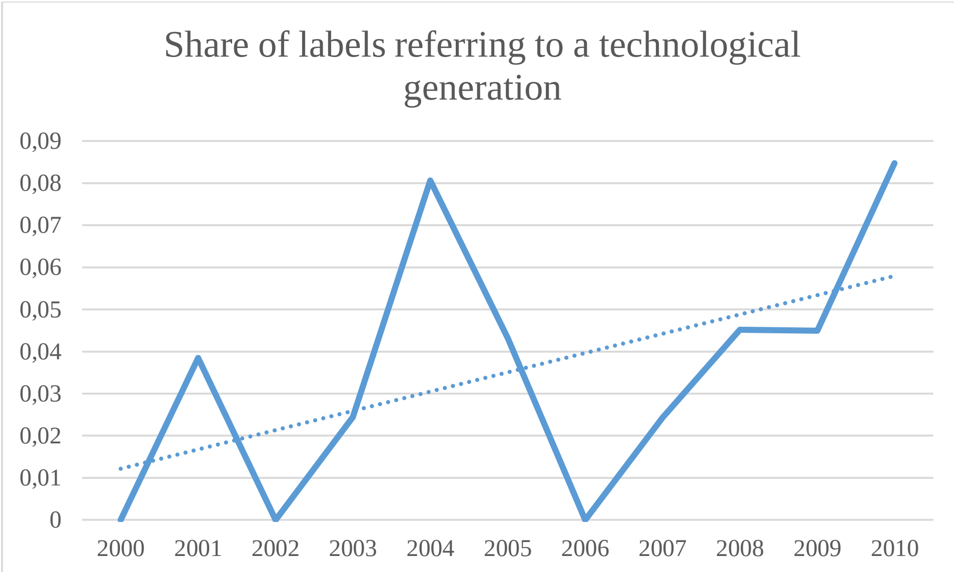
- FIGURE 5: Head nouns per year



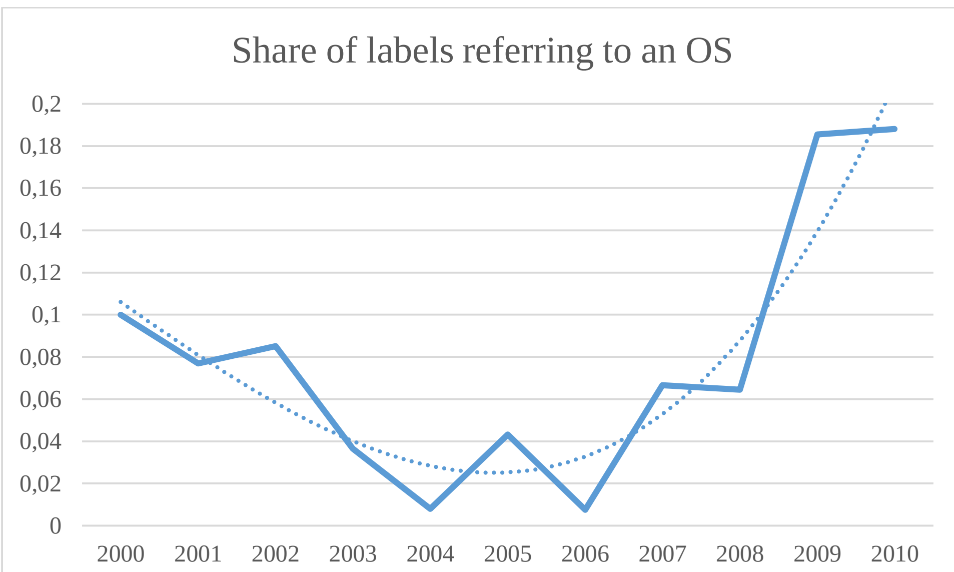
- FIGURE 6: Reference to general technology



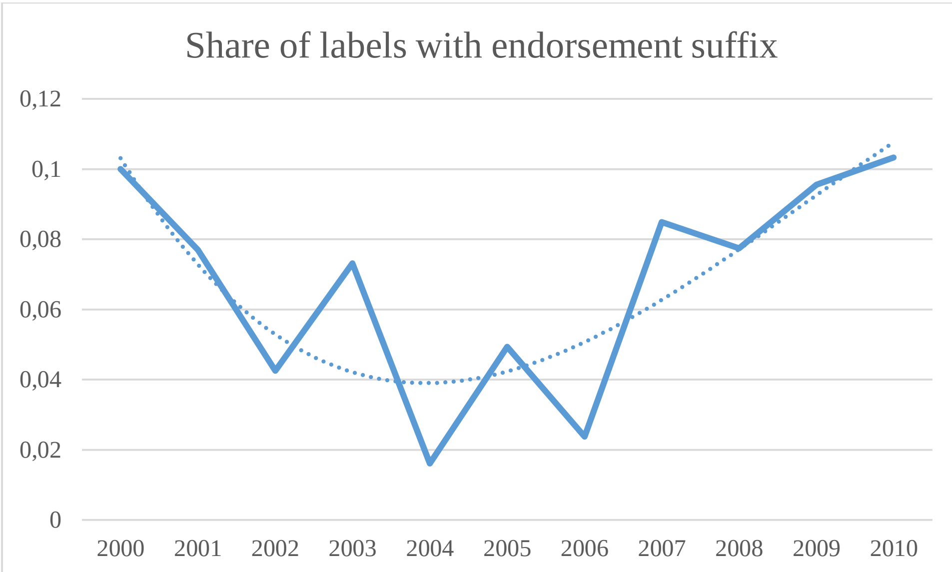
* FIGURE 7: Reference to technological generation



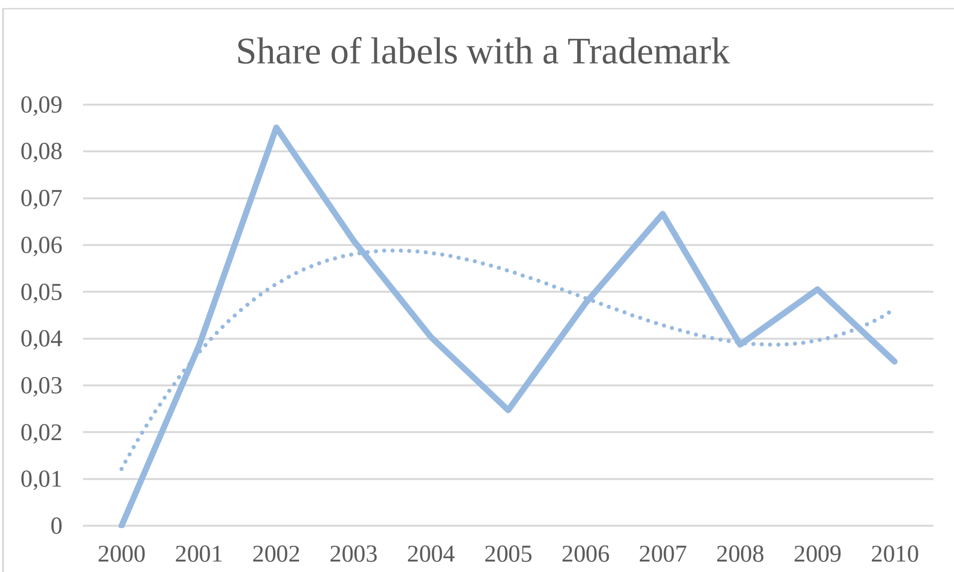
* FIGURE 8: Reference to OS



- FIGURE 9: Endorsement suffix



- FIGURE 10: Share of Labels with a Trademark



- FIGURE 11:Segment Reference

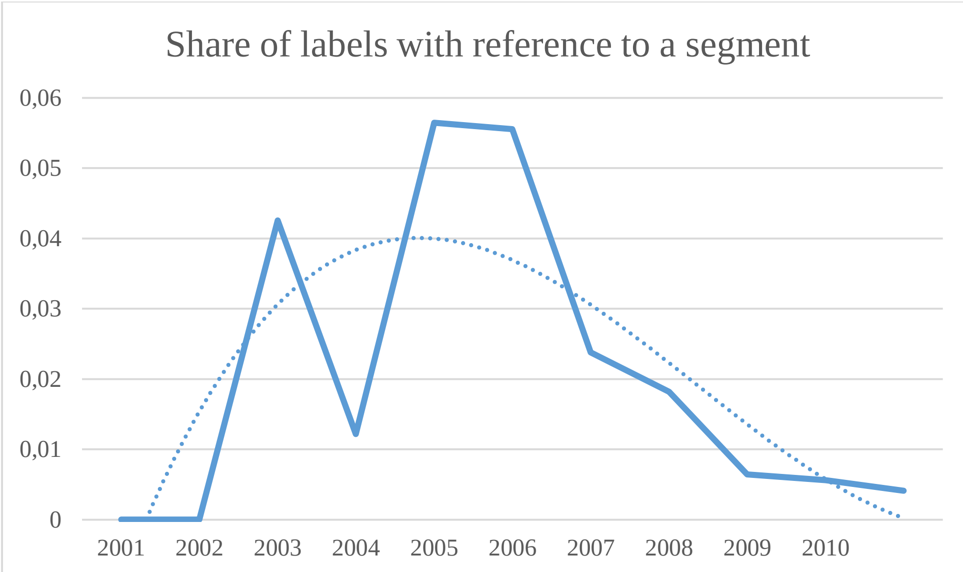


Table 1: Descriptive statistics of the sample and the matched sample

	Unmatched Sample			Matched Sample		
	S.U.	No S.U.	t-test	S.U.	No S.U.	t-test
# of words	2.84	3.70	16.77***	3.14	3.14	0.00
Age	3.55	3.43	0.90	3.63	3.62	0.01
Tech Generation	0.09	0.11	1.98**	0.07	0.07	0.00
Tech Reference	0.43	0.68	11.83***	0.36	0.36	0.00
OS Reference	0.14	0.14	0.17	0.05	0.05	0.00
Trademark	0.07	0.06	0.86	0.03	0.03	0.00
Endorsement suffix	0.13	0.14	0.74	0.03	0.03	0.00
Lifestyle	0.06	0.01	4.93***	0	0	.
Derivation	0.08	0.03	4.94***	0.09	0.03	3.07***
Observations	1571	799		278	278	

Notes: 1:1 CEM procedure matching on "number of words", "age", "tech generation", "tech reference", "OS reference", "Trademark", "endorsement suffix", "lifestyle", "derivation", and "year".

Table 2: the effect of using spatially unified compounds

	(1)	(2)	(3)	(4)
DV:	Log(count)	Log(count)	Log(count)	Log(count)
Method:	OLS	RE	CEM	Alt CEM
S.U. Compound	0.051	0.065	0.415^{**}	0.306^{*}
	(0.05)	(0.04)	(0.13)	(0.12)
Age	-0.102^{***}	-0.097^{***}	-0.037	-0.091^{**}
	(0.02)	(0.01)	(0.03)	(0.02)
Age^2	0.010^{***}	0.009^{***}	0.003	0.009^{***}
	(0.00)	(0.00)	(0.00)	(0.00)
Tech Generation	0.051	0.021	0.209^{**}	0.614^{**}
	(0.05)	(0.05)	(0.06)	(0.14)
Tech Reference	-0.268^{***}	-0.182^{***}	-0.374^{***}	-0.497^{**}
	(0.06)	(0.05)	(0.08)	(0.10)
OS Reference	0.061^{*}	0.113^{**}	0.467^{*}	0.578^{**}
	(0.03)	(0.04)	(0.18)	(0.09)
Trademark	-0.052	-0.009	-0.193	
	(0.03)	(0.08)	(0.12)	
Endorsement suffix	0.041	-0.006	-0.009	-0.100
	(0.05)	(0.05)	(0.10)	(0.07)
Lifestyle	-0.175^{*}	-0.193^{***}		
	(0.08)	(0.05)		
Derivation	0.198^{**}	0.156	0.188^{**}	0.079
	(0.07)	(0.09)	(0.06)	(0.10)
Constant	0.768^{***}	0.733^{***}	0.618^{***}	0.910^{***}
	(0.03)	(0.02)	(0.07)	(0.08)
Head Dummies	Y	Y	Y	Y
Year dummies	Y	Y	Y	Y
N	2191	2191	496	226
R-Squared	0.191	0.169	0.309	0.377

Notes: Head noun clustered standard errors in parenthesis. Year dummies range from 2000 to 2010. Coarsened exact matching based on control variables, Alternative CEM based on control variables and head dummies. Significance levels $^{+}$ $p<0.1$ * $p<0.05$ ** $p<0.01$ *** $p<0.001$