Chapter 4

Marketing research

4.1 Introduction

The current chapter provides an overview of the materials and methods most frequently used in marketing research, with particular reference to those connected with textual data, and reviews selected marketing and consumer studies where content analysis of data is performed. The studies have been selected because of their similarities with the materials and methods used in my preliminary experiments and/or in the final design of the work. Finally, Section 4.3 describes two preliminary experiments to the current work, outlines some theoretical and procedural features common to cultural studies, corpus linguistics and marketing research, and explains how these conflate into the current project.

Quoting from Hair, Bush, and Ortinau (2009, p. 4), marketing research is "the function that links an organisation to its market through the gathering of information".[1] This is a broad definition that encompasses several types of data gathering and analytical activities aimed at providing decision makers with information that might help them plan future action and interaction with the desired audience.[2]

As regards data gathering, data collection is carried out on very many different types of sources. A major important distinction is between primary information, i.e. "information specifically collected for a current research problem or opportunity" (*ibid.*, p. 37), and secondary information, i.e. "information previously collected for some other problem or issue" (*ibid.*, p. 37).

As regards analytical activities, a distinction can be made between qualitative and quantitative approaches and also between exploratory research, descriptive research, and causal research. Qualitative approaches, which involve a limited number of subjects (as few as 8-10 subjects), are fast and inexpensive, but their results can hardly be generalised. For this reason, they are typically adopted in exploratory studies. Quantitative studies, on the other hand, involve a large number of people and are usually performed by means of specifically-made multiple-choice questionnaires.

---

[1] *Marketing research* is not to be confused with *market research*, as the latter focuses specifically on the size and trends of a market and is one of the many faces of marketing research.

[2] As such, marketing research is neither good nor bad in itself. The use that decision makers make of the information gathered, though, may be targeted to gaining personal advantage (as in private business advertisements) or to higher and 'friendlier' goals, as is the case with ethical and social advertising campaigns.

Quantitative studies provide results which can be generalised and are generally performed in descriptive and causal research. Unfortunately, collecting primary data of this type requires careful planning and is highly expensive and time-consuming. In between the two extremes stand rare large scale qualitative studies, and a vast number of quantitative studies performed on a limited number of subjects (as few as 30) for exploratory purposes.

4.1.1 Data gathering in marketing research

Primary information is based on elicited data gathered through a range of direct or indirect questioning techniques (Hair, Bush, & Ortinau, 2009). Direct techniques, such as in-depth interviews and focus groups, involve questioning a small number of subjects on a specific topic and provide the researcher with textual data which is typically analysed using qualitative techniques. The responses collected using direct questioning frequently portray rational and conscious thoughts, as well as socially desirable attitudes (*ibid.*). Indirect techniques, also called projective techniques, include free word association, picture tests, sentence completion tests, and role-playing. Projective techniques – originally developed in the field of psychology – offer a view of the respondent's true opinions and beliefs more neatly than direct ones, and are usually adopted in qualitative studies (Donoghue, 2000). Among the projective techniques used in marketing research, two seem to be of particular relevance in the current research: free word association, and sentence completion tests. Free word association – i.e. "a projective technique in which the subject is presented with a list of words or short phrases, one at a time, and asked to respond with the first thoughts or word that comes to mind" (Hair, Bush, and Ortinau, 2009, p. 185) is among the 10 most common methods used to investigate consumers' needs (van Kleef, Trijp, & Luning, 2005) and has been used, for example, to assess consumers' perception of products (see Guerrero, Claret, Verbeke *et al.*, 2010; Roininen, Arvola, & Lähteenmäki, 2006; Ares, Giménez, & Gámbaro, 2008; Ares & Deliza, 2010) and to assess the cognitive structure of bilingual consumers (Luna & Peracchio, 2002). Interestingly, this is also the technique that anthropologists Szalay and Maday (1973) adopted to study subjective culture, and that Fleischer (2002) used to assess the level of conventionalisation of the image of drinks in Poland, France and Germany (see Chapter 2).

Sentence completion tests – tasks in which "the subjects are given a set of incomplete sentences and asked to complete them in their own words" (Hair, Bush, & Ortinau, 2009, p. 186) have been used, for example, by Belk (1985) to explore the role of materialism in purchase and consumption experiences. In the field of linguistic and cultural studies, this technique was applied by Wilson and Mudraya (2006) to analyse the relationship that exists between naming of shoes and the establishing of different types of shoes as cultural symbols in Russia, and by Potash, de Fileo Crespo, Patel, and Ceravolo (1990) to compare American and Brazilian college students as regards their orientations about future, achievement motives and work ethics, interracial tolerance, and sexuality (see Chapter 2).

Secondary information, i.e. information not specifically collected for the study at hand, includes: customer-volunteered information from electronic customer councils, customer usability labs, e-mail comments, chat sessions, and the like; "data collected by the individual company for accounting purposes or marketing activity reports" (Hair, Bush & Ortinau, 2009, pp. 114-115); or data collected by outside agencies, associations or periodicals. Growing emphasis has recently been put on secondary data, partly as a consequence of the development of the Internet (*ibid.*, 2009, p. 37), and Internet work seems to be gradually replacing field work. Furthermore, secondary information is not as costly as primary information.

4.1.2 Research design in marketing research

Marketing research can be divided into three types: exploratory research; descriptive research; and causal research (Hair, Bush & Ortinau, 2009). Exploratory research aims to outline problems, clarify concepts, collect information, eliminate impractical ideas, and formulate hypotheses. At this stage of research, the researcher can use flexible research designs and methods, and it is customary to resort to convenience sampling, given that the researcher's interest is getting an inexpensive approximation to a specific topic (Guerrero, Claret, Verbeke *et al.*, 2010).[3] Descriptive research is more rigid than exploratory research and requires careful data collection and study design. Descriptive studies can be longitudinal (diachronic) or cross-sectional (synchronic) and aim to describe specific elements of interest, such as the users of a product or service, its demand, the ways it is used, and to make predictions. Finally, causal research performs laboratory and field experiments in order to assess cause-effect relationships between variables (Hair, Bush & Ortinau, 2009).

As regards the analytical methods used in marketing research, these vary depending on the type of data and study. As Aggarwal, Vaidyanathan and Venkatesh (2009) point out, one of the first analytical methods applied to textual data in the marketing field was content analysis; since its first use in the late '70s, it has been adopted, for example, to analyse advertisements, to determine the knowledge structure of salespeople, and to understand communication on home shopping networks. In marketing research, content analysis seems to be preferably applied in exploratory rather than descriptive studies, possibly due to the problems associated with sampling and measurement, or the reliability and validity of content categories, as well as the prohibitive cost of manually coding large amount of data.

Content analysis has been variously defined. Neuendorf (2002, p. 10) lists some of the definitions offered by 'main players in the development of quantitative message analysis'; the elements common to all those definitions suggest to describe content analysis as a quantitative analysis of textual messages (of any type) by means of systematic and replicable measures. As Weber (1990, p. 12) clarifies, "a central idea in content analysis is that the many words of the text are classified into much fewer content categories". Content analysis categories can be decided *a priori*, or while analysing the data. In either case, finalising the coding scheme requires several review steps that go hand in hand with application of the coding scheme to different sets of data by different coders. As we shall see in Section 4.2, although definition of the coding scheme before looking at the data is strongly advocated by content analysis

[3] According to Graveter and Forzano (2008, cited in Guerrero, Claret, Verbeke *et al.*, 2010), convenience sampling is probably used more often than any other kind of sampling in behavioural science research.

guidebooks, such as Weber (1990) and Neuendorf (2002), establishing categories while looking at the data seems to be the preferred option by most researchers in the marketing field. Finally, as Weber (1990) and Neuendorf (2002) clarify, central issues in content analysis are reliability and validity of the classification procedure. Reliability is guaranteed by a consistent application of the coding scheme. When coding is performed manually, different coders should be able to code the same text in the same way: this can be attained by creating and using of a specific codebook which describes the coding categories and explicates how the codes should be interpreted. Automatic coding, on the other hand, requires the use of specific software based on dictionaries and can lead to a higher degrees of consistency. Validity refers to the extent to which the categories adopted in the analysis represent or measure the concept that the researcher is interested in. It is interesting to notice at this point that automatic content analysis is largely applied in other disciplines that are linked to culture, such as the social sciences (see for e.g. McTavish & Pirro, 1990), or linguistics and cultural studies (e.g. Wilson & Moudraia, 2006).

A type of automatic content analysis which has recently been undergoing significant development and is finding application in marketing research is sentiment analysis, or opinion mining,[4] in the form of assessment of positive, negative or neutral sentiment in text. Sentiment analysis has gained momentum with the development of the Web 2.0 – rich in opinionated text types, such as blogs, review portals and other user-generated contents – and by the application of computerized text mining, information retrieval and natural language processing procedures to secondary data available on the Web (Tsytsarau & Palpanas, 2012). Sentiment analysis, which is ultimately performed electronically, often starts with manual analysis of small text samples which are then used to train the specific software (*ibid.*, p. 484). Indeed, determining sentiment polarity is a highly context-sensitive task (Choi, Kim, & Myaeng, 2009). Sentiment analysis may be applied at document, sentence, clause, or even word/phrase level depending on the type of text and the research goals (Thet, Na, Khoo, & Shakthikumar, 2009). Since consumers are largely relying on on-line opinions when making their purchasing decisions (Kaiser, Schlick, & Bodendorf, 2011; Archak, Ghose, & Ipeirotis, 2011), user-generated product reviews (and sometimes also blogs, weblogs and message boards) are analysed in order to understand the standing of a given product on the Web (see for example the articles listed in the Literature Survey section in Jebaseeli & Kirubakaran, 2012). This, however, is not the only possible application of sentiment analysis in marketing. Other applications include, for example, deriving the pricing power of a product feature (Archak, Ghose, & Ipeirotis, 2011) and warning marketing managers about the rising of critical situations (Kaiser, Schlick, & Bodendorf, 2011). Sentiment analysis is worth mentioning in the current work because it analyses positive/negative polarity of text with a logic which is similar to that of semantic prosody (see Chapter 3, Section 3.6.4), and also because of its use of Web text. However, the tools and methods

[4] The two terms are generally used interchangeably, although they originated in different communities and consequently have slightly different notions. As Tsytsarau & Palpanas (2012) explain: “Opinion Mining originates from the IR [Information Retrieval] community, and aims at extracting and further processing users’ opinions about products, movies, or other entities. Sentiment Analysis, on the other hand, was initially formulated as the NLP [Natural Language Processing] task of retrieval of sentiments expressed in texts. Nevertheless, these two problems are similar in their essence [...].”

adopted in sentiment analysis and the applications that have been made of it go beyond the scope of the current work. For this reason, no specific sentiment analysis paper will be reviewed in the next section.

Finally, in marketing studies, as in many other scientific fields, it is common practice to validate the results of qualitative studies using quantitative data, and *vice versa*, or different analytical techniques. Yu, Shen, Kelly and Hunter (2006) validated the results obtained from a questionnaire-based quantitative study on 51 subjects by comparing them to the findings of a focus-group meeting. Guerrero, Claret, Verbeke *et al.* (2010) checked the robustness of results obtained with the semantic analysis of free word association tasks by comparing them to the findings of focus-group discussion.

The following section reviews a few marketing and consumer studies where content analysis of data is performed. The studies have been selected because of their similarities with the materials and methods used in my preliminary experiments and/or in the final design of the work.

4.2 Review of selected marketing studies

Content analysis techniques are typically applied to elicited data and used in small-scale studies.

Ares, Giménez and Gámbaro (2008) analysed free word associations of the images of five types of natural yogurt. Fifty Uruguayan subjects were asked to evaluate the images and write down the first thoughts that came to their minds. The associations thus elicited were semantically analysed: for each type of yogurt, terms with similar meaning were manually grouped into categories and the categories shared by less than 10% of the participants were discarded; next, the categories observed for the different yogurts were further classified into 19 final categories. For each category, word frequencies were counted and used to compare the different types of yogurt to each other. The results showed that regular yogurt was considered a healthy product having pleasant texture and flavour; low-calorie yogurts were mainly associated with diet or slimming and with texture or other type of sensory defects; finally, yogurts enriched with fibre and antioxidants were mainly related to health, and the prevention of diseases. As the authors declare (*ibid.*, p. 641), “word association thus provided an interest insight into consumers’ perception of yogurts, which could be useful for product development and marketing”.

Codern, Pla, de Ormijana, and Gonzales (2010) employed content analysis to identify the dimensions that lay people and healthcare professionals use to assess the risk of smoking. To this purpose they carried out focus-group interviews with 11 users and 7 professionals. The focus-group discussions were transcribed and manually coded by the researchers. The coding system was developed in subsequent stages: full reading of the transcripts and identification of the recurring topics; review of the codes; and code categorisation into groups. More concretely,

> “two researchers individually generated codes and categories that were then contrasted in search for differences and commonalities. A third researcher followed the process, reading the transcriptions and verifying the codes and their meanings. The three

> researchers involved in the analytical process met regularly to discuss emerging issues." (*ibid*. 2010, p. 1565)

This is a perfect description of the steps and processes used in the current work to create the initial version of the coding scheme. In some preliminary experiments, two separate coders went through texts about chocolate in English and in Italian and identified the recurring semantic fields; the categories thus separately established were then compared and contrasted in order to create a single list of codes which was reviewed by a third coder (myself). The three coders met frequently to discuss coding issues.[5]

Guerrero, Claret, Verbeke *et al.* (2010) used free word associations of the node word *traditional* to assess the perception of traditional food products in six European regions (Flanders in Belgium, Burgundy in France, Lazio in Italy, the counties of Akershus and Østfold in Norway, Mazovia in Poland and Catalonia in Spain). About 120 subjects in each region were individually asked to name three words in response to the verbal stimulus word *traditional*, while concentrating on food-related issues. For each region, gender and age group, frequencies of elicitation were obtained at three different levels: first, at the level of the words elicited; second, by classifying the elicited words in 55 semantic classes; third, by grouping the 55 classes in ten principal dimensions. Analysis at the level of the 55 semantic categories showed a general tendency of southern European regions to associate the idea of *traditional* with broad concepts such as Heritage, Culture or History, while central and northern European regions tended to focus more on practical issues such as Convenience, Health and Appropriateness. Analysis at the level of the ten principle dimensions showed fewer differences between geographical regions, but highlighted gender differences: women seemed to prefer the Heritage, Health, Origin or Sensory dimensions, while men the Elaboration, Habit, Marketing and Variety ones. Finally, the authors compared results of the word association study to the results of focus group interviews, which confirmed their robustness. This study by Guerrero *et al.* is similar to my experiments in its using a double level of semantic analysis (a wider number of semantic categories, subsequently grouped into a smaller number of broader semantic domains).

The studies reported above applied content analysis to free word associations. However, content analysis is frequently applied to focus groups transcripts and open-ended questions. For example, Brug, Debie, van Assema and Weijts (1995) carried out an explorative study on people's motivation in consuming fruit and vegetables. Data were collected in focus group interviews. The focus group transcripts were analysed by dividing the sentences into groups depending on their content, each group representing a specific issue which emerged during the discussions. Finally, the issues thus identified and grouped were used to prepare summaries of the focus group meetings.

More interesting is a study on critical success factors in construction project briefing by Yu, Shen, Kelly and Hunter (2006). The authors submitted a questionnaire to 51 experienced construction practitioners. Alongside background information, the questionnaire included an open-ended question aimed to collect opinions on the

---

[5] A detailed description of the genesis of the coding scheme, along with it subsequent revisions is included in the Appendix.

success factors of project briefing. The open-ended question responses were analysed by assigning responses to coded categories. Through this procedure, 37 critical success factors were identified and coded; the critical success factors were subsequently grouped into five major categories adapted from a careful study of the scientific literature on the topic. Finally, the results of the open-ended responses were compared to the results of a focus group meeting, which confirmed their validity.

In very recent times, however, some researchers seem to be experimenting the application of manual or automatic content analysis on non-elicited data and on what could be considered secondary information.

Aggarwal, Vaidyanathan and Venkatesh (2009), used lexical analysis of the semantic Web to assess the positioning of different brands relative to that of competitors.[6] By means of Google API, they searched the Web for sentences containing specific brand names (e.g., "Stetson") and analysed their co-occurrence with selected adjectives and descriptors (e.g., "up-to-date"), considering such co-occurrence an indication of subjectivity, i.e. a subjective evaluation or opinion. Significant co-occurrence was established by means of the mutual information score. Finally, frequency of co-occurrence (over a vast amount of textual data) was used to infer brand's positioning. This study is highly interesting to the current work primarily because of its use of the Web as source of data. Second, because it considers frequency of occurrence of semantic associations as an indication of shared opinion among several subjects.

Finally, a study by Kleij and Munsters (2003) is worth mentioning here, though it does not apply content analysis procedures. The authors involved 165 subjects in the evaluation of different varieties of mayonnaise. The participants were asked to specify their preferences for each type on a 10-point liking scale. Furthermore, they were given the option to freely comment on their assessments. The words in the freely expressed comments were analysed in terms of word co-occurrences. Finally, word co-occurrences were counted for each different product, and the relationship between products and product characteristics as verbalised by the respondents were visualised be means of correspondence analysis. The results of the analyses were compared to preference mapping, a standard procedure in the analysis of sensory drivers of liking based on the use of objective data from trained panel assessment of product characteristics. The authors reported that "the agreement between the correspondence map and the preference map is striking, with the additional advantage being that the correspondence map is stated in terms of consumer language" (*ibid.*, p. 43). This study is relevant to the current research because of its using analytical methods typical of corpus linguistics: first, words are counted (by producing a frequency word list);[7] second, for each of the most frequent words (e.g. *taste*), co-occurrence (collocation) with other words is considered and discussed (e.g. *taste – sour*). Third, because these corpus linguistics analytical procedures are offered as an innovative methodological approach complementary to more traditional preference mapping.

[6] For a discussion of the Web as corpus, see Chapter 3.
[7] The authors of this study do not use this term, but clearly the words counts they mention correspond to a frequency word list.

4.3. Cultural studies, corpus linguistics, marketing research, and the current work: common features

As we have mentioned, the current work aims to assess the possibility of using materials and methods typical of corpus linguistics for an analysis of cultural associations of a given node word which could find theoretical or practical applications not only in the linguistic and cultural fields, but also in the marketing one. Such a type of analysis should bank on some common ground among the three fields. For the current purposes, I have identified the necessary common ground in the following features: word associations; semantic/content analysis; and frequency as a measure of the association's importance.

Word associations appear in their psychological dimension in free word association and sentence completion tasks, and in their linguistic dimension in text in general and collocations in particular. Indeed, some parallelism can be seen between empirical collocations and verbal associations or EMUs, to use Szalay and Maday's 1973 terminology, see Chapter 2).[8] Empirical collocations are words that co-occur in the same textual environment; and frequency of co-occurrence determines collocational strength. The collocates of a node word, once grouped into semantic fields or domains, show its semantic preference (Partington, 2004). Analogously, EMUs co-occur in the same psychological environment as the word that triggers them, and they all show high collocational strength to the node word. Classification of words/sentences into semantic/thematic categories is the basic principle of content analysis.

Finally, a higher frequency of one EMU over another could thus be an indication of a cultural (vs. an individual) origin of the EMU itself. This last observation may be better understood considering Fleischer's (1998) theory of culture, illustrated in Chapter 2, according to which the cut-off line between individual and cultural mental associations is frequency of appearance across different subjects belonging to the same cultural group. Furthermore, frequency of elicitation of words in free word association tasks has been related with the strength or importance of a concept in the consumers' minds (Guerrero, Colomer, Guàrdia, Xicola, & Clotet, 2000).

As regards the textual material to use, elicited data in the form of sentence completion tasks or free sentence writing – widely used source of intelligence in marketing research – seems to be accepted, though is not the preferred type of data, in corpus linguistics at least when it comes to analysing culture (see for e.g. Fleischer, 2002; and Wilson and Mudraya, 2006).[9] Certainly, they are in keeping with the definition of corpus I subscribed to in Chapter 3. On the other hand, the use of large

[8] Interestingly, some recent empirical research has shown "a direct predictive relationship between the statistics of word co-occurrence in text and the neural activation associated with thinking about word meanings" (Mitchell, Shinkareva, Carlson, Chang, Malave, Mason, & Just, 2008, p. 1191; Murphy, Baroni, & Poesio, 2009). These results suggest that a direct relation between co-occurrence of words in text and the mental lexicon may exist, though further research is needed in this field.

[9] The term 'elicited data' has been frequently frowned upon by corpus linguists, because it is connected to introspection, a practice that according to some "does not give evidence about usage. […] Actual usage plays a very minor role in one's consciousness of language and one would be recording largely ideas about language rather than facts of it" (Sinclair, 1991, p. 39). This, however, is not a generalised view (see for example Fillmore, 1992; and Nordquist, 2009).

general textual data is more common in corpus linguistics than in marketing. Finally, in both disciplines, the Web is a relatively recent, but promising source use textual data. Consequently, the current work will use data elicited through sentence completion and free sentence writing tasks as a sort of 'control' situation to which Web data can be safely compared.

4.3.1 Preliminary experiments

A preliminary experiment in the analysis of EMUs using corpora of non-elicited data was attempted by Bianchi (2007). The study aimed to highlight EMUs to chocolate in contemporary Italian society and compare the analytical possibilities offered by general and specialised corpora in a task of this kind. Concordances were generated for the Italian words for chocolate in a specialised corpus about chocolate and in a general-purpose corpus. Each concordance line was manually classified in terms of semantic context of the node word, that is the main topic(s) mentioned in the relevant text segment. Classification was based on the lexical meaning of the co-text and was performed through a data-driven, open-coding system. Semantic contexts were then grouped into higher-order categories, which were called 'conceptual fields'. Comparison between the two corpora highlighted what appear to be long-existing and well-established EMUs for chocolate in Italian society. It also suggested the possibility of evolution in the psychological associations of chocolate from the 1980s to 2005. From a methodological perspective, the findings seemed to show that suitable data for cultural analysis can equally be retrieved from a very large general corpus, or a small-to-medium-sized specialised corpus, provided that they include a wide variety of texts by different authors, and that in cultural analysis, the major concern in corpus creation, along with text variety, seems to be time-coverage. In terms of analytical methods, the two levels of analysis used – conceptual fields (higher-level; less fine-grained) and semantic contexts (lower-level; more fine-grained) – were both highly, but differently useful: conceptual fields helped establishing that, despite their apparent differences, these corpora could be considered samples from the same population, and guided the researcher in making sense of results and in establishing some kind of ranking between groups of psychological associations; semantic contexts, on the other hand, was the level where the most interesting EMUs emerged.

Another preliminary experiment (Bianchi, 2010) investigated the suitability of different methodological approaches to automatic semantic tagging in the analysis of cultural traits as they emerge from subjective meaning reactions to given words (EMUs). A first goal of this study was to compare the potential of manual coding to automatic tagging. To this aim, two sets of data elicited from British native speakers were coded manually as well as with Wmatrix, an automatic semantic tagger (see Chapter 5), and for each set of data the results were compared at the level of conceptual domains (superordinate, broader categories) and of semantic fields (subordinate, more fine-grained categories). In order to compare manual and automatic tagging, a specific conversion scheme was developed and applied. At the level of conceptual domains, the conversion scheme was applied to the top 30 items in the semantic frequency list and in the semantic keyword list of the elicited data as offered by Wmatrix, excluding grammatical items. As an intermediate step between manual tagging (sentence-based) and semantic tagging (word-based), it was decided to consider also the top 30 items of the raw frequency list and of the keyword list, as this allowed manual tagging to be applied on the basis of individual words. Therefore, the top 30 semantic items in the lists (excluding the node word) were manually mapped to one or more of the conceptual domains used in manual coding. Those analyses were then compared to the results of manual coding of the whole elicited datasets, which showed that the semantic frequency list performed generally better than the other lists. In fact, it retrieved the same or a higher number of domains and systematically showed strong correlation values at the Spearman test. At the level of semantic fields, comparison was performed using the most frequent 50 items in the semantic frequency list and in the semantic keyword list. When using the semantic frequency lists, the data consistently showed levels of correlation in the modest range; when using the semantic keyword list, results were less consistent. Another goal of the study was to compare elicited data to Web data. At this level of analysis, comparison was performed using automatic tagging only. Consequently, 10,000 sentences were extracted from a general Web corpus for each of the node words of the elicited data and the Web datasets thus created were tagged with Wmatrix. The semantic word lists of the Web data were compared to the semantic word lists of the elicited data. For the sake of experimentation, correlation was computed in three different ways: (1) using the whole semantic frequency lists, (2) using the top 100 items in the lists; and (3) using the top 50 items. All the six cases (three for each node word) showed interesting positive correlation between the elicited and the Web data, the strength of the correlation decreasing from strong to medium to low-medium as the number of items considered decreased.

4.3.2 The current work

Banking on the preliminary experiments described above and on the theoretical ideas reviewed in the previous chapters, the current work will use elicited data gathered through free sentence-completion and sentence-writing tests. The data elicited will be analysed following a content analysis procedure highly similar to that described in Codern, Pla, de Ormijana, and Gonzales (2010) and results will be discussed within the framework of cultural systems theories. Furthermore, the results obtained with elicited data will be compared to non-elicited data from a Web corpus. Indeed, if (freely available) Web corpora gave the same results as more traditional marketing research techniques, marketing research could benefit from a wider range of fast and inexpensive methods. Finally, an automatic semantic tagger will be tested on the elicited data, in order to assess the extent of its possible application in cultural analysis.

The materials and methods employed in the current research are described in detail in Chapter 5. The various analyses performed on the elicited data and on the Web data are reported in Chapters 6-10.