

Linguistic support for concept selection decisions

J. DELIN,^{1,2} S. SHAROFF,² S. LILLFORD,³ AND C. BARNES³

¹Enterprise IG, London, United Kingdom

²Centre for Translation Studies, School of Modern Languages and Cultures, University of Leeds, Leeds, United Kingdom

³School of Mechanical Engineering, University of Leeds, Leeds, United Kingdom

(RECEIVED March 20, 2006; ACCEPTED November 25, 2006)

Abstract

Affective engineering is being increasingly used to describe a systematic approach to the analysis of consumer reactions to candidate designs. It has evolved from Kansei engineering, which has reported improvements in products such as cars, electronics, and food. The method includes a semantic differential experiment rating candidate designs against bipolar adjectives (e.g., attractive–not attractive, traditional–not traditional). The results are statistically analyzed to identify correlations between design features and consumer reactions to inform future product developments. A number of key challenges emerge from this process. Clearly, suitable designs must be available to cover all design possibilities. However, it is also paramount that the best adjectives are used to reflect the judgments that participants might want to make. The current adjective selection process is unsystematic, and could potentially miss key concepts. Poor adjective choices can result in problems such as misinterpretation of an experimental question, clustering of results around a particular response, and participants' confusion from unfamiliar adjectives that can be difficult to consider in the required context (e.g., is this wristwatch “oppressive”?). This paper describes an artificial intelligence supported process that ensures adjectives with appropriate levels of precision and recall are developed and presented to participants (and thus addressing problems above) in an affective engineering study in the context of branded consumer goods. We illustrate our description of the entire concept expansion and reduction process by means of an industrial case study in which participants were asked to evaluate different designs of packaging for a laundry product. The paper concludes by describing the important advantages that can be gained by the new approach in comparison with previous approaches to the selection of consumer focused adjectives.

Keywords: Adjective Rules; Affective Engineering; Concept Evaluation; Kansei Engineering; Semantic Differential Experiment

1. INTRODUCTION

It is important for economic success that companies have a robust and effective product development process to drive innovation and deliver new products to market. This is even more critical in the mature and competitive fast moving consumer goods (FMCG) market, as it is well known that over 80% of new products fail. Many different approaches are used to generate new products, the most successful of which use consumer insights and opinions to evaluate prototypes and ideas. This clearly is important in this market where functionality and ergonomics are market entry char-

acteristics, and demanding consumers desire appealing products that complement lifestyles and aspirations.

However, many of the current consumer-based methodologies are founded upon qualitative techniques and require significant subjective interpretation to turn the data into product designs. Affective engineering is gaining credibility as a method to evaluate candidate designs against consumer perceptions to deliver products that elicit positive consumer reactions and thus increase desire and purchase intent.

The field has its roots in Kansei engineering (Nagamachi, 1995), which has reported improving products as diverse as footwear (Solves et al., 2006) and cars (Nagamachi, 1999). The core of the method is a semantic differential experiment involving demographically selected participants who are asked to rate candidate designs against a series of bipolar

Reprint requests to: Catherine Barnes, School of Mechanical Engineering, University of Leeds, Leeds LS2 9JT, UK. E-mail: c.j.barnes@leeds.ac.uk

adjectives (e.g., attractive–not attractive, traditional–not traditional). The results of the experiment are analyzed statistically to identify the design features that most correlate with the positive consumer reactions. These knowledge-based rules can then inform future product development.

A number of key challenges emerge from this process. Clearly, it is vital that suitable designs are available to span the breadth of design possibilities. However, it is also paramount that the most appropriate adjectives are used to describe the product and its desired brand identity accurately, and to reflect the judgments that participants might want to make. Unsuitable adjective choices can result in a range of problems, including the following:

- participants misinterpreting an experimental question resulting either in a flat distribution of responses or “double peaks” that evidence two separate interpretations;
- adjectives with very similar semantic meanings artificially causing results to cluster around a particular response, leading to too much weight being accorded to a single feature; and
- participants being either “led” or confused by adjectives that are unfamiliar to them, or difficult to consider in the required context (e.g., is this wristwatch “oppressive”?).

Kansei engineering suggests that adjectives are selected either by talking to potential consumers and designers or by searching trade press and relevant literature (Nagamachi, 1995). These approaches are relatively unsystematic, and rely upon trial and error to avoid problems; thus, words may be missed that represent key parameters for informing product design. This issue is especially relevant when designing FMCG products and their packaging as the influence of brand is a key aspect of the process and no method currently exists to ensure that this is represented in the adjective set.

The aim of this paper is to present a method to ensure that adjectives with appropriate levels of precision and recall are presented to participants in an affective engineering study in the context of branded consumer goods; this is described in Section 2. The computations involved in defining, extending, and pruning the list of adjectives are presented in Section 2.2. Section 3 introduces the manual application of linguistic rules to further filter to find the most suitable adjectival set. Finally, in Section 4, an illus-

trative case study is presented to show how this process can be used with an affective engineering process to support the concept selection decision within a product development process.

1.1. Affective engineering support for product development

Analysis has shown (Childs et al., 2006) that most large companies run similar product development processes, which are broadly reflective of classical engineering design theories. However, this work also showed that the boundaries between the development stages are not common. Figure 1 shows this common process although it does fail to demonstrate the iteration within and between stages.

Although the analysis showed that there was general agreement on the product development process, the tasks carried out (or knowledge required) by the different companies at the different stages of pack development varied, depending on the strategic goal of the development process. These can be categorized as the following:

- creating a new market, with a step change innovation;
- targeting a gap in the market;
- extending an existing brand, with a new product; and
- refreshment of an existing brand and product.

Currently the voice of the consumer is integrated into this process largely using qualitative research techniques such as focus groups and observations; however, there are issues over both the reliability of this research and how it is interpreted by the development team, leading to costly development iterations and potentially market failures.

It is important that companies listen to their consumers in a structured manner and design their products to appeal, and this requires the following:

- a multidisciplinary team with skills in consumer understanding (lifestyle preferences, etc.), functional requirements (product containment, etc.), and design to creatively link the consumer needs to pack capabilities in an appealing way;
- a process for consumer testing that can quantitatively elicit needs to feedback into the design process;
- opportunity for iteration to ensure that the concept is fully in line with consumers needs; and

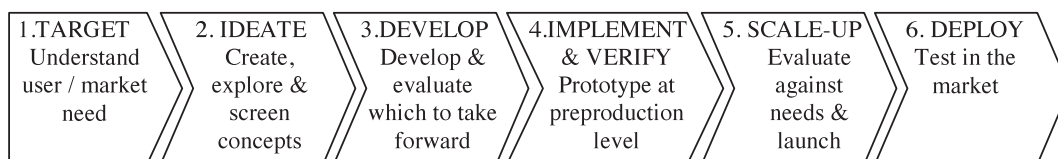


Fig. 1. A generic industrial product development process.

- evaluation criteria and methods to measure success against criteria, previous concepts, or competitors.

Affective engineering supports communication across the disciplinary boundaries of the team and informs the conceptual design decision-making process when compromising between appeal and function. It helps to do the following:

- make transparent the consumer's impression of the concept and their underlying requirements,
- make fewer avoidable iterations because of clearer interpretation of consumer data and resultant design guidelines,
- make decision making more informed, and
- confirm that the product and brand qualities are being communicated through the concept.

2. IDENTIFICATION OF RELEVANT ADJECTIVE SET

2.1. Manual selection of appropriate seed words

The first step in producing a list of adjectives reflecting a product's brand identity is to establish the initial field for exploration. An FMCG product can be described in terms of the triad of *functional qualities* of the product itself, its *pack*, and the *brand equity*. For instance, in terms of a product's functional qualities, people buying a washing powder consider characteristics such as cleaning clothes, perfume, retaining shape and color, and so forth. In terms of the pack, washing powder comes in boxes (usually cardboard), which have different important qualities, such as its size and shape and, how easy it is to grasp, and so forth. A branded washing powder can be described through the system of values that brand owners attribute to potential buyers, such as *caring for the family*, or in terms of the product's "personality," such as *engaging*, *genuine*.

Original lists of functional qualities and packaging words can come from general knowledge about the domain or from market research, which identifies a list of benefits or characteristics that consumers treat as being important qualities of a product when buying and using it. The source for building the list of brand-specific concepts for a branded product is a strategic branding statement that outlines the company's vision of how the product can be made to appeal to consumers, and how they perceive tangible and intangible characteristics of the product.

The resulting set of principles that are used to evaluate products therefore derive from a triad of characteristics, deriving from the pack, the overall function of the product, and the desired brand position. These characteristics may be captured in a single word, a phrase, or even a paragraph (from "soft" to "ensures no static cling"). From these characteristics, "seed words" [a word that is used as a basis for the remainder of the process and forms the key input into the British National Corpus (BNC)] must be derived to begin

the process of developing a set of adjectives for evaluation of the product.

Each member of the function–package–brand triad produces a list of several (usually 5–15) concepts, not all of which are immediately useful for a list of seed words, as some of them can be specialized terminology, descriptive phrases, or unusual words, which are not frequent in the general language. The next step involves exploration of each concept to find words touching several aspects in a domain identified by the concept. For instance, the functional property of "perfume" can give rise to such words as *scent*, *odor*, *aromatic*, and *deodorized*. The properties of the pack can give rise to *handle*, *hold*, *steady*, and so forth, whereas the brand values can produce *care*, *engaging*, *love*, and so forth.

The resulting list of seed words consists of about 50–70 words, and captures the company's perception of qualities that people consider as a basic requirement of the total product that they buy and the differentiator that makes one product better than the other for their purposes. However, the list of seed words is not ready for affective engineering experiments, as it suffers from problems both in precision and recall. From the viewpoint of recall, there can be many other expressions, some of which are more suitable. From the viewpoint of precision, the list of seed words can include concepts with inherent ambiguity, which may not immediately be clear to the participant.

2.2. Automated extension of seed words list

In this step we improve both recall and precision of the seed word list by comparing it against examples of real language use. The source for language data in the current study is the BNC, which is a 100 million word collection of texts from a wide range of sources, designed to represent a comprehensive picture of how British English is used (Aston & Burnard, 1998). Recall can be improved by extending the list with other words that are used in similar contexts, whereas precision can be improved by taking words that stand out in the original list as they do not share many contexts with others. The procedure for improving recall is completely automatic, whereas precision can be improved by manually pruning the list of candidates following a set of rules.

Automatic detection of similar words is based on the distributional similarity hypothesis (Harris, 1985), according to which two words have similar meaning, if they share a significant number of other words occurring in their context. There have been several proposals for computing lists of words with similar meanings. They can be classified into two groups. One approach involves detection of typical patterns that can relate words, such as AND/OR, IS-A, and so forth (Pantel & Ravichandran, 2004): if words frequently co-occur in similar patterns (e.g., *frustration and anger*), they have similar meanings. The second approach involves representing the context of each word by a context vector (also known as feature vector):

$$C(w) = \{w_1, w_2, \dots, w_n\}.$$

Features in this vector are weights, which can represent collocates (other content words co-occurring in the window of N words), syntactic features, such as subject, object, or modifier relationships, or the strength of association of word w with documents in which it occurs. In our case, the feature space was represented by a list of collocates ranked according to the log-likelihood score (Manning & Schütze, 1999), which in comparison to other measures (such as chi-square or mutual information) ensures that both common expressions (e.g., *enjoy life* or *spend one's life*) and less frequent terminological constructions (*life jacket* or *life imprisonment*) receive high weights. Then the procedure involves computing the distance between vectors representing respective words $D(C(w_1), C(w_2))$ and finding words most similar in the feature space. The distance can be measured in several metrics, such as cosine, Jackard, Dice, and so forth (see Manning & Schütze, 1999, chap. 8.5).

The advantage of the first (pattern-based) approach is that it can be efficiently applied to large-scale dynamically changing corpora, because it is based on a relatively small number of topical patterns. However, it requires either development of elaborate parsers to apply patterns to running text or an additional step for extraction and evaluation of suitable patterns. The second (context-vector) approach produces high-quality lists automatically, but it is computationally expensive, because most typically the dimensionality of the vector space is huge. For instance, in our case we wanted to collect statistics on the most typical environment for every sufficiently frequent content word, so we computed collocates for words where frequency is above 50 occurrences in 100 million words of the BNC. This leaves us with the feature space of about 35,000 items (examples of words at the bottom of our frequency list are *imprudent*, *tarragon*, and *uprating*).

Given that computing the distance between vectors of 35,000 dimensions is computationally expensive, the next step in the automatic procedure was to reduce the dimensionality of the feature space by means of singular value decomposition (SVD) of the resulting matrix of collocates following the procedure designed by (Rapp, 2004). SVD-based transforms were also used in latent semantic analysis (LSA) to compute the similarity between terms in documents (Landauer et al., 1998). However, in our case we do not deal with terminology, so our matrix is based not on the co-occurrence of words per document, but on the strength of collocations between words in their contexts.

SVD transform involves decomposition of the initial matrix of word co-occurrence $M = U\Sigma V^T$, in which Σ is a diagonal matrix containing singular values of the original matrix M (Berry et al., 1999). The selection of k largest singular values of Σ gives a reduced dimensionality space, in which the original $n \times n$ matrix of correlation between content words can be approximated by $n \times k$ matrix, in which vectors for n words (35,000 in our case) contain only

k features (300 in our experiment). This reduces the dimensionality substantially without losing information about the relationship between words.

For each word in the original seed list we produce its simclass consisting of about 20 words, which SVD-reduced vectors have the closest distance to the vector of the source word according to the cosine metrics. For instance, *love* produces the following simclass:

1. *lover* (0.367), *passion* (0.366), *god* (0.361), *never* (0.353), *loving* (0.350), *mother* (0.334), *life* (0.319), *heart* (0.319), *beautiful* (0.318), *affection* (0.318), *hate* (0.310), *husband* (0.307), *desire* (0.304), *passionate* (0.295), *kiss* (0.294), *pleasure* (0.289), *friendship* (0.287), *friend* (0.280), *father* (0.275).

Not all words in the list are suitable for affective engineering, but recall here is more important than precision, as the initial investigation could have missed words like *beautiful* or *passionate*.

At this stage we deal with the issue of precision by first using collocate lists to remove seed words, which frequently co-occur with words outside of the desired domain, and second, by checking the ambiguity of respective simclasses. For instance, even if *care* is suitable for the original list of brand keys, the set of its collocates reveals that its most typical contexts are *intensive care*, *residential care*, and *home care services*, which can cause problems with interpretation of these words by participants in affective experiments. This suggests that *care* is not suitable for an affective engineering test, as it is likely that it in the mind of some subjects it will be linked to health care, and hence, result in misinterpretations.

WordNet (Fellbaum, 1998) is a lexical database that also captures the relationship of synonymy between lexical items and is frequently used for research in computational linguistics; however, its coverage of affective words is not good enough. For instance, its output for *care* included words such as *attention*, *aid*, *tending*, *caution*, and *guardianship*, but did not highlight the subtlety of the health-care analogy, an aspect critical for affective engineering studies. This is because it uses a thorough but thesaurus-based approach rather than considering how a word can be used in natural language.

The ambiguity from the BNC output can produce two unrelated strands in their simclasses, such as the simclass of *spirit* (because of corpus processing, words are in lowercase):

2. *god* (0.453), *jesus* (0.424), *holy* (0.403), *divine* (0.389), *faith* (0.377), *soul* (0.366), *christ* (0.350), *cider* (0.339), *whisky* (0.330), *sin* (0.328), *grace* (0.326), *heaven* (0.326), *sherry* (0.323), *wine* (0.315), *apostle* (0.311), *lager* (0.310), *prophet* (0.307), *demon* (0.307), *blessing* (0.302).

At the same time simclasses of other words, which belong to the same semantic field as *spirit*, but are less ambiguous, such as *resilient*, can produce interesting simclasses:

3. *energetic* (0.395), *resistant* (0.343), *durable* (0.341), *robust* (0.331), *timid* (0.326), *lovable* (0.312), *tough* (0.295), *tolerant* (0.293), *unpalatable* (0.291), *adaptable* (0.286), *remarkably* (0.281), *prone* (0.277), *minded* (0.274), *docile* (0.273), *sturdy* (0.272), *willed* (0.272), *talented* (0.267).

In our experiments we also found that packaging words, for example, *bottle*, typically referred to the content of the pack and not to its inherent qualities. The simclass for *bottle* in the BNC is the following:

4. *beer* (0.631), *wine* (0.623), *whisky* (0.611), *champagne* (0.556), *gin* (0.554), *brandy* (0.548), *glass* (0.540), *sherry* (0.536), *vodka* (0.536), *pint* (0.532), *jar* (0.524), *lager* (0.511), *drink* (0.508), *jug* (0.500), *cider* (0.469), *flask* (0.468), *mug* (0.461), *scotch* (0.460), *rum* (0.456).

The most frequent collocates of *bottle* also refer to the content (most typically beverages): *drink half a bottle of brandy*, *buy a bottle of Scotch*, *an empty vodka bottle*, with occasional *milk* or *hot-water bottles*. In 100 million words of the BNC there are only seven examples of *bottle* combined with *shampoo*, none of which really evaluates qualities of the pack. This confirms the position of branding agencies that “the pack communicates the product and the brand,” so no evaluative words could be generated from the packaging description as such.

In further studies we concentrate on product and brand seed words exclusively. In both cases the aim is to generate

an exhaustive range of language: the wider the range of language the better. The range of words that are used to describe the product creates its *semantic space* (see Fig. 2). However, as examples 1–4 show, the semantic space for a product contains various words, many of which are not suitable for an affective engineering test, for example, *lover* or *never* for *love*, whereas others are relevant, but are nouns or verbs, whereas in our tests we use adjectives. What is more, we also need gradable adjectives for which one can apply a judgment on a scale: a plastic box cannot be judged as more or less plastic. Thus, in the next step we convert the semantic space to produce gradable adjectives that can occur in constructions like *to be M X*, where *M* stands for a modifier, for example, *more*, *very*, *not so*. For instance, we can convert *pleasure* from example 1 into *pleasant*, *hate* into *hateful*, *friend* into *friendly*, but we cannot produce suitable adjectives for *kiss* or *mother*.

3. ADJECTIVE REDUCTION PROCESS

The next task is to reduce the list back down to the 10 or 20 best candidates that can be presented to participants in semantic differential experiments. Given that the process as described so far can generate several thousand words, the accuracy of the reduction process is crucial: it is important to ensure that the more insightful and useful words remain in the list, whereas those that are less descriptive of important qualities of the pack, product function, or brand are removed. It is also vital to remove words that are obscure or may be unknown to participants, that have more than one relevant interpretation, or that are simply difficult to apply to products of the nature of that being tested. Any such words would introduce error and noise into the data produced from this process.

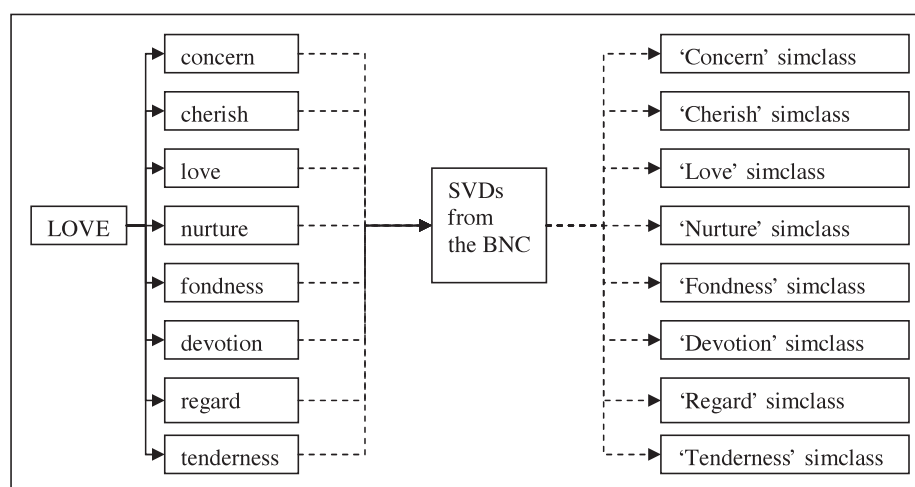


Fig. 2. An example of semantic space.

3.1. Manual application of expert rules

As yet, no automatic process is available to reduce the burden of the task to remove unsuitable words from the generated list. This manual process uses the application of a set of linguistically and grammatically informed rules, examples of which are given in Table 1. This is a relatively easy process whereby each rule in turn is applied to the list of words and any adjective that violates the rule is removed.

After these rules have been applied, the adjective list is considerably shorter, and many of the words that are not central, or that are likely to produce confusion or error, have been removed.

The next step reduces the list still further by reapplying the remaining words back to the set of keys resulting from the brand, product, and pack description. In this step, we find that some of the words satisfy more than one such characteristic, which makes them strong candidates for inclusion in the test. This step may, in some cases, be the final one, as the words that are selected may be only those that satisfy more than one quality or criterion. In practice, however, it is also necessary to ensure that all the characteristics have one or more words that relate to them included in the test. Thus, the selection process should include not only words that relate to more than one criterion, but also some words that may relate to only one, but that are its sole representative in the test.

3.2. Adjectival candidates for consumer survey

All adjectives that claim more than one “parent” quality are good candidates for use in the tests. However, further pro-

cedures are sometimes necessary to reduce a still overlarge word list down to the 10 or 20 adjectives that are manageable for experimental participants. This step is done in collaboration with the brand owner, who has ultimate oversight of the final list. For example, although in our case study (see Section 4) the adjectives “tender,” “luxurious,” and “conventional” occurred in the context of more than one brand quality, the brand owner did not favor them and felt that other candidates were better. Although this is a subjective step, it also relates to the fact that brand descriptions are not themselves perfect and do not always represent fully the character of the brand. In practice, the more specific and distinctive a brand description is, the more distinctive and representative the words that this process generates will be. A more general brand description is likely to generate less interesting and incisive words and less valuable results for the brand owner.

4. ILLUSTRATIVE CASE STUDY

We illustrate our description of the entire adjective expansion and reduction process by means of an industrial case study, completed in February 2006 for a commercial client, in which participants were required to evaluate different designs of packaging for a cleaning product. Client confidentiality requires that only some representative words and results could be reported in this case study.

The aim of the study was twofold:

1. to investigate how the client’s existing product packaging was perceived by customers in relation to other similar packs currently on the market; and

Table 1. *Example of rules used to reduce adjective set*

Rule I	Remove adjectives that are not plausibly related to objects.	Although “friendly,” for example, may be metaphorically extended to apply to inanimates, words such as “enthusiastic” and “unbiased” are not.
Rule II	Remove adjectives that describe purely evaluative reactions.	For example, “good” and “nice” do not describe products, but people’s feelings about them, and feelings of liking and disliking, do not inform us about specific product-related reactions.
Rule III	Remove ambiguous adjectives.	For example, “clear” might be a good word to refer to either a concrete or an abstract quality of products that promote cleanliness (such as clear skin) or that are themselves clear (as in without color), but it is frequently used to describe the clarity of an argument or a conclusion.
...		
Rule XI	Remove adjectives that describe feelings.	For example, “passionate” is a word that frequently occurs in brand descriptions, but the pack or product itself cannot be passionate.
...		
Rule XIII	Remove words relating to prolonged experience with the sample rather than a brief experimental encounter.	Respondents cannot be expected to know whether a product will be “satisfying” or “durable,” because it relates to experiences of use, rather than quick evaluation.

Table 2. *Extending seed words*

Brand Equity		Product Benefits	
BE ₁ delight	Enjoyment, surprise, joy, content, glad, happy, pleasure, delighted, etc.	PB ₁ skin kindness	Kind, caring, nonirritating, smooth, moisturizing, etc.
BE ₂ high standard	Best, exceptional, exclusive, extraordinary, unique, etc.	PB ₂ perfume	Scent, odor, aroma, etc.
BE ₃		PB ₃	
BE ₄		PB ₄	
...		...	
BE _n		PB _n	

2. to validate that the resultant adjectives from the generation and reduction process give robust and meaningful results, as defined by the client.

As expected, this description will focus on the results that relate to the second objective.

4.1. Identification and reduction of relevant adjective set

The first stage in the process was to identify a set of appropriate seed words for input into the system. The client, a linguistic expert, and an affective engineering expert defined a set of seed words under the headings of functional benefits, brand equity, and packaging attributes. In this situation, packaging attributes did not result in any suitable adjectives, so it will not be discussed any further within this case study. These seed words were manually extended using the artificial intelligence (AI)-supported linguistic process described in Section 2.2 to about 70 words (see Table 2). The list of extended seed words were in agreement with the client to ensure they accurately reflected the product and brand essence.

Each seed word was then input into the BNC as described in Section 2.2 to automatically identify other adjectives that have similar lexical behavior in naturally occurring language. A list was produced of the most significant collocations for each sufficiently frequent word from 100 million words of the BNC. The SVD method was used to group words that occur in similar lexical contexts.

Each seed word resulted in 10–20 significantly related words from the SVD method, representing the saturation of the semantic exploration for the product and brand attributes. Table 3 shows part of the relevant adjective set placed into columns under the heading of their original seed word. The numbers indicate the cosine semantic distance between the vectors of respective words in the resulted SVD matrix; this indicates the similarity between the derivative and original word, but is not used to discriminate in this methodology.

The next stage was to reduce the size of the relevant adjective set into the range of adjectives suitable for the

affective engineering consumer study. This is achieved using the rule set hierarchy and accompanying grammatical tests as described in Section 3. Table 4 shows a representative selection of adjectives that were removed along with the rule violation reasons.

Table 3. *Part of the relevant adjective set*

Delight		Love	
Delight		Love	
Joy	−0.56	Lover	−0.37
Pleasure	−0.56	God	−0.36
Beauty	−0.34	Never	−0.35
Astonishment	−0.33	Loving	−0.35
Dismay	−0.33	Mother	−0.33
Delighted	−0.33	Life	−0.32
Enjoyment	−0.32	Beautiful	−0.32
Enjoy	−0.32	Affection	−0.32
Passion	−0.32	Hate	−0.31
Amusement	−0.3	Passionate	−0.3
Surprise	−0.3	Pleasure	−0.29
Enjoyment		Concern	
Pleasure	−0.49	Concerned	−0.46
Satisfaction	−0.38	Interest	−0.38
Enjoy	−0.36	Problem	−0.38
Joy	−0.33	Worry	−0.36
Fun	−0.32	Anxiety	−0.35
Excitement	−0.31	Sympathy	−0.33
Amusement	−0.31	Importance	−0.32
Pursuit	−0.29	Question	−0.32
Enrich	−0.27	Fear	−0.3
Thrill	−0.27	Controversy	−0.3
Enrichment	−0.26	Implication	−0.28
Surprise		Curiosity	−0.28
Shock	−0.51	Cherish	
Surprising	−0.41	Inseparable	−0.29
Dismay	−0.4	Enshrine	−0.27
Puzzle	−0.38	Perversely	−0.26
Astonishment	−0.38	Ardent	−0.25
Delighted	−0.37	Worst	−0.25
Astonish	−0.37	Companionship	−0.25
Unexpected	−0.37	Undying	−0.25
Wonder	−0.36	Finicky	−0.24
Disappointed	−0.35	Serenity	−0.24

The final adjective list for the consumer survey is highlighted in Figure 3 and included words such as *tender*, *conventional*, *fun*, *luxurious*, *showy*, *everyday*, *slender*, *cosy*, and *bold*.

Three additional adjectives were added to test whether the selection process was discriminatory. These were chosen because they represented concepts that the client was keen to include:

- “*uncomplicated*”; “*satisfying*”—(violation of Rule XIII, see Table 1)
This is related to judgments that associated interacting with a product over an extended period of time, which is clearly outside the scope of the survey. Although it was interesting to find out that “perceived functionality” judgments were able to be made purely from a visual stimulus.
- “*passionate*”—(violation of Rule XI, see Table 1)
This is related to how a user can feel toward something rather than what the object suggests.

4.2. Consumer survey

To validate the process described, the selected adjectives were carried forward into a full consumer survey. Fifty female participants took part in the experiment between the ages of 35 and 50 who were all frequent users of premium products within the market sector and who all considered added value aspects important (e.g., brand promise, smell, and other detailing).

A set of 10 prototype samples was used for the study with all graphics, branding, and color removed to constrain the evaluations to pure shape. Figure 4 shows an exemplar sample used for the study.

Each participant was asked to complete a semantic differential experiment. Figure 5 shows the semantic differential questionnaire as used in this survey, illustrating how



Fig. 4. An example sample stimulus used in the consumer survey.

each adjective was presented with its corresponding negative. Each participant was asked to evaluate each sample, in turn, according to each adjectival pair and rate it on a 7-point scale. Each participant completed one questionnaire per sample and was allowed to do the experiment at his or her own pace.

4.3. Results and discussion

The participants' scores were aggregated and cleaned. The raw evaluation data and preference scores were analyzed primarily with regard to the following points:

1. evaluation and preference score distribution,
2. principal component analysis of the subject's perceptual space, and
3. contribution of adjectival selection method.

4.3.1. Preference score distribution

Preference score distribution plots show the randomness of the participants' scores. If the AI-supported method described in this paper has not resulted in suitable adjectives, and there is little/no significant correlations between the evaluative adjective and the object then one of the following scenarios will be observed for the adjectives (see Fig. 6).

- a. Flat distribution: Responses are equally distributed across the scale showing no overall consensus by the test group of participants.
- b. Central peak distribution: A cluster of responses at the central point showing that most participants had no significant opinion of the interaction between the adjective pair and sample.
- c. Double peak distribution¹: Two polarizing opinions of the group suggesting the evaluative adjective is suitable for differentiating the properties of the shape, but participants have opposing opinions of the correlation.

These distributions are frequently found in such studies and require that the data be removed. However, that was not the case in this experiment, as all the responses for the identified good adjectives showed the following:

1. useful correlations for evaluating objects, and differing degrees of consensus were observed across different adjective and sample combinations, which evidences the ability of participants to use a set of evaluative adjectives to communicate differences in perceptual properties between the samples; and
2. a lowered level of ambiguity in comparison to previous studies.

¹This may not necessarily be an issue, as it evidences two separate populations of opinion and thus another potential market opportunity.

Test item 10		
Tender	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	Not Tender
Not Everyday	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	Everyday
Not Fun	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	Fun
Conventional	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	Not Conventional
Not Luxurious	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	Luxurious
Slender	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	Not Slender
Not

Fig. 5. An example survey questionnaire.

4.3.2. Principal component analysis of the subject's perceptual space

Although the samples can be represented in a perceptual space where each evaluation adjective has an independent dimension, it is too complicated to determine the relative location of each sample in such dimensional semantic space. Moreover, it is not known whether there exists any interaction among the adjectival words.

Therefore, a principal component analysis (with a varimax rotation and using significance as eigenvalues > 1), can define the underlying similarities and relationships between the individual words and samples.

Two components were extracted, which means that there are two orthogonal sets of perceived similarities between the adjectival relationships to the samples: principal component 1 (PC1) accounts for 68% of total variance, and PC2 for 21% of total variance. The two adjectives that load highest onto PC1 are luxurious and showy; cosy and friendly load very highly onto PC2. Figure 7 shows diagrammatically the semantic space with all sample scores plotted. From this one can see that sample 2 loads highly on both PC1 and PC2, making it the ideal sample for the brand if it wants to be seen as both stylish and natural. If one wants a sample that is only perceived as natural then samples 3 or 9 are appropriate. However, if the target is to have a sample that is only stylish, then one should select sample 5. Sample 4 scores poorly against both factors, so is unlikely to represent the right design.

4.3.3. Contribution of adjectival selection method

The principal component analysis highlighted that “satisfying” and “uncomplicated” did not relate significantly with any other underlying response, and further analysis showed that this was explained by the flat response distributions across all participants. As these were included to test the adjective generation rules (they violated the linguistic rule set), the fact that they have been of no use in this experiment goes towards validating the rules.

Passionate was also included as a test of the rule set, but it loaded highly in PC1. A review of the response distribu-

tions for passionate showed that participants were able to make the interpretation of a “not passionate” bottle, but were not able to relate passionate positively to any of the bottles, thus confirming its position as a nondiscriminatory adjective.

5. CONCLUSIONS

Affective engineering has evolved from Kansei engineering developed over the last 20 years in Japan. However, in this time no systematic and repeatable process has been documented for the definition of an appropriate set of adjectives used in the consumer survey. This paper reports an AI-supported method for determining a list of adjectives that has shown, through the illustrative case study, to help to reduce experimental bias, misunderstandings, and confusion during the completion of the semantic differential questionnaire, thus improving accuracy and confidence in the results.

The process brings together the notion of natural word collocations in language usage to ensure extensive coverage of potential adjectives and a linguistically informed rule set to remove inappropriate selections. The resulting adjective set was related back to the original seed words to identify the most suitable subset for use within the consumer survey.

Analysis of the results has shown that the advantage of applying the AI-supported process was that it was able to repeatedly and robustly define a suitable adjective set that provided rational results (as defined by the client). The data distributions showed that the participants were able to complete the questionnaire with little confusion or misinterpretation. In addition, it was shown that the adjectives introduced to test the sensitivity of the process all showed responses characteristics that the AI method was shown to reduce.

Further testing of this process on real industrial studies is required to ensure the suitability of the method in all situations.

If necessary, other large text collections can be used, including domain-specific ones, such as journalistic texts

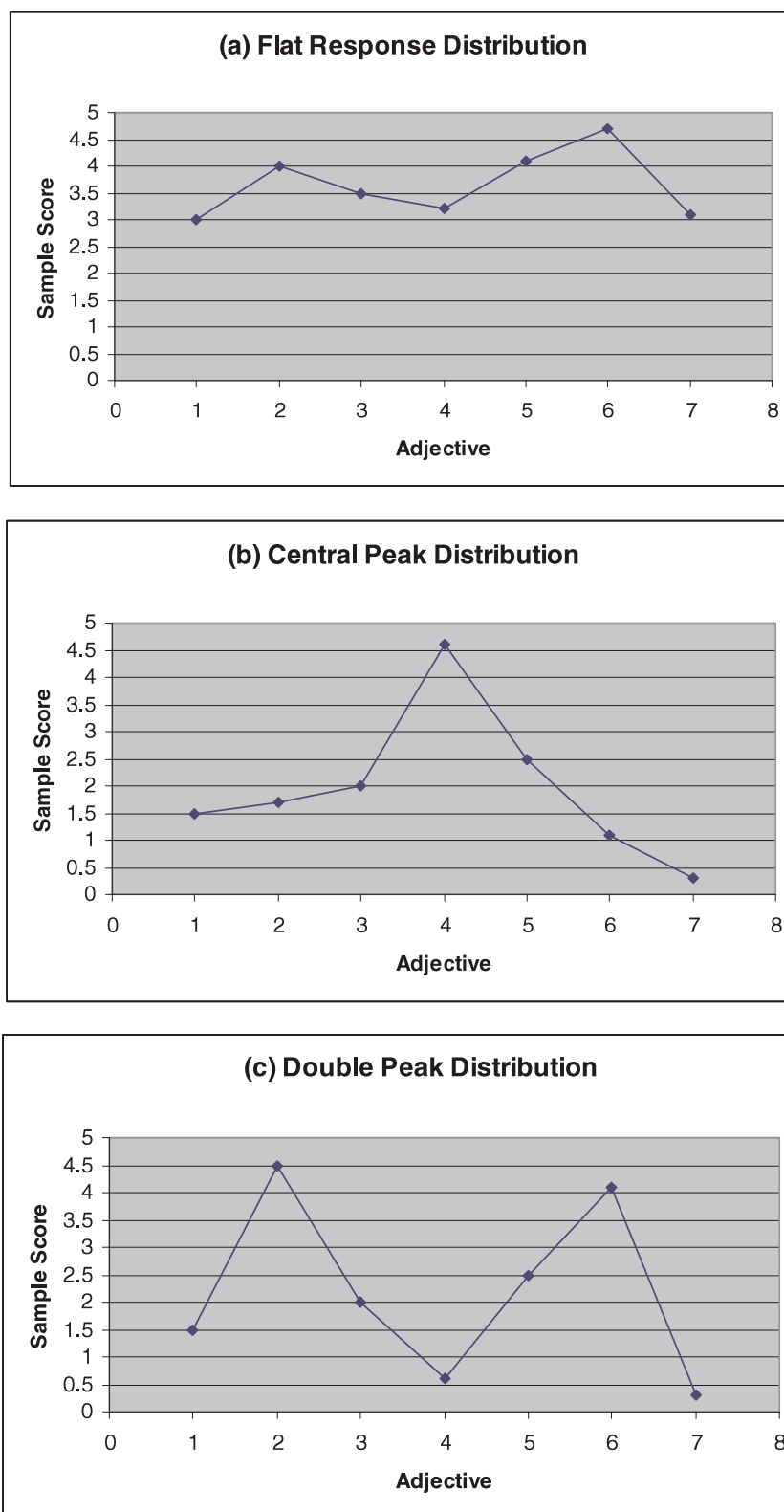


Fig. 6. Examples of response distributions. [A color version of this figure can be viewed online at www.journals.cambridge.org]

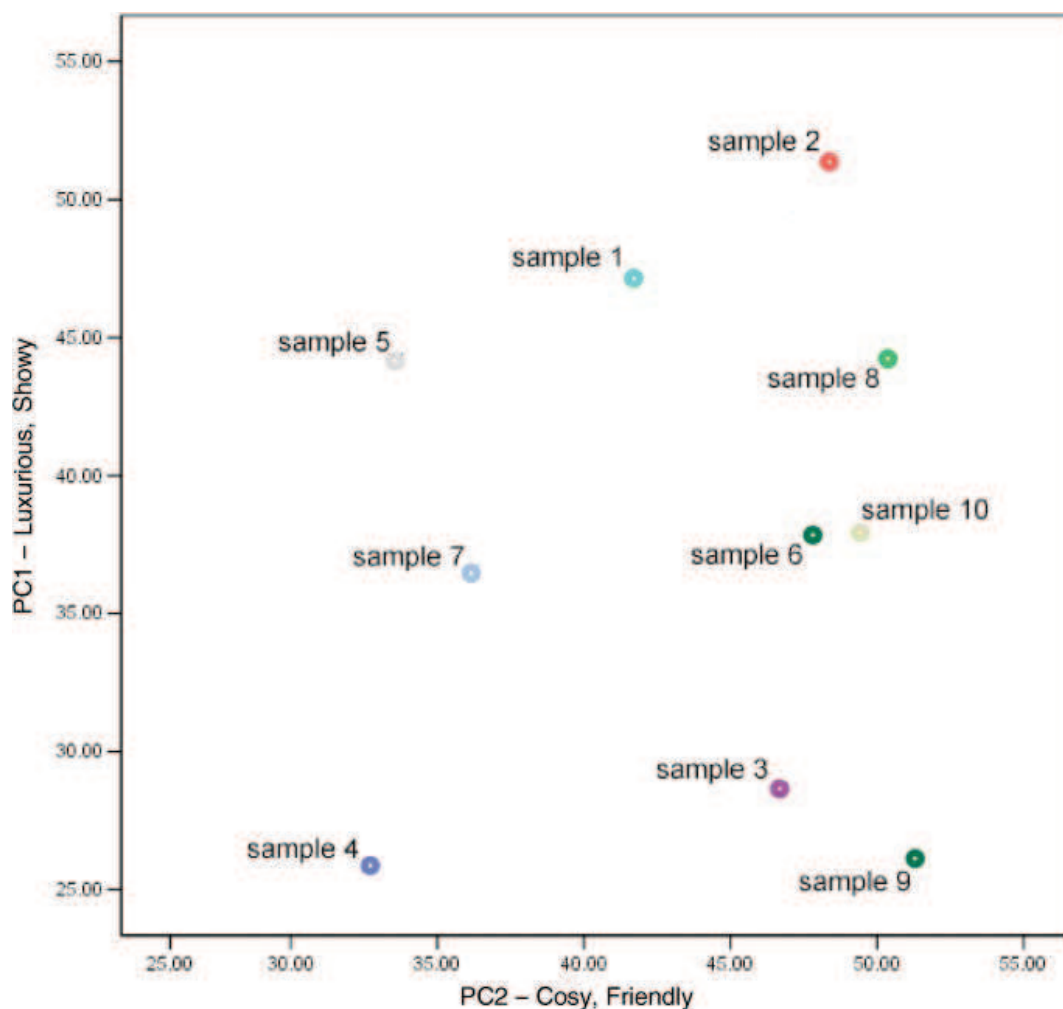


Fig. 7. A semantic map showing sample ratings versus principal components. [A color version of this figure can be viewed online at www.journals.cambridge.org]

or evaluative language from blogs. The disadvantage of such collections is that they may not cover the language the respondents are exposed to in their daily lives, but the advantage is that include more words and collocations on a specific topic. The methodology is independent of language, so if there is a sufficiently large text collection in another language, such as a Web-derived representative corpus (Sharoff, 2006), it can be also used to do a study in another language and to compare word lists and customers' perception. This approach is better than straightforward translation of adjectives into other languages, as semantic meanings can often differ widely in other countries. This clearly is an issue in the FMCG domain where the future lies in global branding.

ACKNOWLEDGMENTS

The work reported in this paper was carried out as part of Knowledge Transfer Partnership Number 4445 in conjunction with Dr.

W. Lewis of Faraday Packaging Partnership and M. Hancock of PIRA International Ltd. We thank the participating company members of the Faraday Packaging Partnership, who had significant input into the definition of this process. Special thanks to the client company who provided the case study reported here.

REFERENCES

- Aston, G., & Burnard, L. (1998). *The BNC Handbook: Exploring the British National Corpus With SARA*. Edinburgh: Edinburgh University Press.
- Berry, M., Drma, X., & Jessup, E. (1999). Matrices, vector spaces, and information retrieval. *SIAM Review* 41(2), 335–362.
- Childs, T., Agouridas, V., Barnes, C., & Henson, B. (2006). *Controlled appeal product design: a life cycle role for affective (Kansei) engineering*. Engage Network work package 2. Accessed at www.engage-design.org
- Fellbaum, C. (1998). *WordNet. Electronic Lexical Database*. Cambridge, MA: MIT Press.
- Harris, Z. (1985). Distributional structure. In *The Philosophy of Linguistics* (Katz, J.J., Ed.), pp. 26–47. New York: Oxford University Press.
- Landauer, T.K., Foltz, P.W., & Laham, D. (1998). Introduction to latent semantic analysis. *Discourse Processes* 25, 259–284.

- Manning, C., & Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*. Cambridge, MA: MIT Press.
- Nagamachi, M. (1995). Kansei engineering: a new ergonomic consumer-oriented technology for product development. *International Journal of Industrial Ergonomics* 15, 3–11.
- Nagamachi, M. (1999). *Kansei Engineering and Its Applications in Automotive Design*, SAE Technical Paper 1999-01-1265.
- Pantel, P., & Ravichandran, D. (2004). Automatically labeling semantic classes. *Proc. HLT/NAACL-04*, pp. 321–328.
- Rapp, R. (2004). A freely available automatically generated thesaurus of related words. *Proc. 4th Language Resources and Evaluation Conf.*, pp. 395–398.
- Sharoff, S. (2006). Open-source corpora: using the net to fish for linguistic data. *International Journal of Corpus Linguistics* 11(4), 435–462.
- Solves, C., Such, M.-J., Gonzalez, J.C., Pearce, K., Bouchard, C., Gutierrez, J.M., Prat, J., & Cruz Garcia, A. (2006). Validation study of Kansei engineering methodology in footwear design. In *Contemporary Ergonomics 2006* (Bust, P.D., Ed.), pp. 164–168. London: Taylor & Francis.

Judy Delin is Visiting Professor of Language and Communication in the Centre for Translation Studies, University of Leeds, and Head of Brand Language, Enterprise IG (EIG), a WPP company. Judy has a BA in English studies from the University of Nottingham (1984) and a PhD in cognitive science and linguistics from the University of Edinburgh (1990). Dr. Delin has published extensively on linguistics of English and other languages. At EIG, Judy leads a team that helps companies articulate brand position, and researches relationships between language, personality, and consumer reaction. Major clients include telecoms, insurance, energy, and consumer goods companies.

Cathy Barnes is a Lecturer in design and manufacture integration in the School of Mechanical Engineering at the University of Leeds and a Network Manager at Faraday

Packaging Partnership. Her research interests focus on the human interface of design and manufacturing, and include affective design, emotional tribology, and decision-based concurrent engineering; she has published extensively in these areas. Dr. Barnes is leading the development of affective engineering tools in a funded collaboration with nine major consumer goods companies, and has particular experience of experimental design, especially self-report elicitation of user feelings about products.

Stephen Lillford is a Knowledge Transfer Partnership Associate working on a project to transfer Kansei engineering tools into the packaging industry. He has a BEng (Honors) in mechanical engineering from the University of Leeds (2002) and an MA in design and branding strategy from Brunel University (2003). He has worked on a number of research programs working across public and private partnerships developing tools for industrial use.

Serge Sharoff is a Lecturer at the Centre for Translation Studies, University of Leeds. Dr. Sharoff's research interests are related to three domains: linguistics, cognitive science, and communication studies. He has published extensively in this area. Probably the most interesting part of his recent research is semiautomatic acquisition of representative corpora from the Web. The current set of resources includes 100–200 million word corpora for Chinese, English, French, German, Italian, Polish, Portuguese, Russian, and Spanish. Another recent development is ASSIST, a joint project with Lancaster University, which is about an automatic procedure for finding translation equivalents between English and Russian.