ABSTRACT: Flavor lexicons are a widely used tool for documenting and describing sensory perception of a selected food. Development of a representative flavor lexicon requires several steps, including appropriate product frame of reference collection, language generation, and designation of definitions and references before a final descriptor list can be determined. Once developed, flavor lexicons can be used to record and define product flavor, compare products, and determine storage stability, as well as interface with consumer liking and acceptability and chemical flavor data.

Lexicon characteristics
A flavor lexicon is simply a set of words to describe the flavor of a product or commodity. A lexicon is then applied or practiced using descriptive sensory analysis techniques. A lexicon, like a specific technical dictionary, provides a source list to describe a category of products, such as commodities (cheese, beef, fish) or finished products (bread, orange juice, coffee, chocolate). Although the descriptive panel generates its own list to describe the product array under study, a lexicon provides a source of possible terms with references and definitions for clarification (see ASTM lexicon comments).

As a broad example, the ASTM Lexicon (Civille and Lyon 1996) provides hundreds of words that might be selected by a descriptive panel to describe any number of product categories. The clear admonition for use of the ASTM Lexicon and/or for generation of any lexicon is to:
1. collect a product frame of reference
2. generate the terms
3. review references and examples
4. develop a final descriptor list

When a panel begins work in a product category, the 1st source of terminology is the wide sample set or frame of reference. The panel needs to evaluate several (25 to 100) products in the category that cover the widest array of flavor characteristics. The Aromatics, Tastes, and Feeling Factors that the samples manifest are the basis for the lexicon. After some initial attempts at organizing the terms generated by all of the panelists, the panel and panel leader organize references to provide clarification of the grouped terms. Often many terms relate to original ingredients and the processes they are subjected to in the creation of the category. Sensory acuity, as well as for interest in sensory panel work. Following extensive training (approximately 60 h) panelists work with the group to generate the language and the method for sample presentation and evaluation. Panelists are screened for sensory acuity, as well as for interest in sensory panel work. Following extensive training (approximately 60 h) panelists work with the group to generate a consensus profile of the sensory properties of the product and the intensity of each. In addition, overall amplitude of the product, as well as a measure of balance and blend are evaluated. The scale used for character (descriptor) notes is a 4 point scale (0-not present, 1-slight, 2-moderate, 3-strong) with half point designations in between, for a total of 7 points of discrimination. Attributes are evaluated in order of appearance. Aftertaste.

Descriptive Analysis Techniques
The 1st published descriptive sensory technique is the Flavor Profile Method (FPM) developed in the 1950s by Arthur D. Little, Inc. (Pangborn 1989; Lawless and Heymann 1998; Meilgaard and others 1999). More recent applications emerged in the 1970s with Quantitative Descriptive Analysis (QDA) and the Spectrum™ method of descriptive analysis. The latter 2 techniques differ markedly from FPM in that they are designed to take measurements from individual panelists and then generate a panel average, rather than generation of a group consensus profile as with FPM (Piggott and others 1998). A detailed overview of all 3 techniques can be found in the Manual on Descriptive Analysis Testing for Sensory Evaluation (1992). Panelist selection is a critical aspect of descriptive analysis (Zook and Wessman 1977; Lawless and Heymann 1998; Meilgaard and others 1999). A good prospective flavor panelist should be able to describe sensations and should be able to discriminate between taste and aroma sensations. Typical screening tests include taste and aroma acuity and discriminatory ability. Good health, enthusiasm, and usage of the general product category are also characteristics of good prospective sensory panelists. All 3 techniques involve screening panelists for sensitivity to flavor and aroma and discriminatory ability followed by extensive training to generate a group of individuals who can function analogously to an instrument to evaluate flavor of products.

In FPM, generally fewer panelists are used, with a minimum of 4 (Keane 1992). The panel leader is a member of the panel and works with the group to generate the language and the method for sample presentation and evaluation. Panelists are screened for sensory acuity, as well as for interest in sensory panel work. Following extensive training (approximately 60 h) panelists work as a group to generate a consensus profile of the sensory properties of the product and the intensity of each. In addition, overall amplitude of the product, as well as a measure of balance and blend are evaluated. The scale used for character (descriptor) notes is a 4 point scale (0-not present, 1-slight, 2-moderate, 3-strong) with half point designations in between, for a total of 7 points of discrimination. Attributes are evaluated in order of appearance. Aftertaste.
may also be evaluated. The group consensus process was not designed for statistical analysis, although the scale may be modified to allow statistical analysis.

In QDA, the panel leader is a sensory professional, rather than a member of the panel. During training, the panel (ideally 8 to 12 prescreened individuals) generates the language to describe the product (Stone 1992). The panel leader does not participate in the discussion to generate the flavor attributes, but facilitates the process. The panel also determines the order of appearance of the attributes and, thus, order of evaluation for each descriptor. Definitions for each descriptor are generated and additional products, which may further clarify terms, may also be used. The panel also practices rating products to become familiar with the process and to gain confidence in their abilities. Data is collected on scorecards using line scales anchored on each end. Panelists mark the scale using a dash. The marks are converted to numerical scores by measuring the responses on the scale with a ruler, digitizer, or computerized system. The data can be analyzed statistically and is traditionally graphically represented using a webplot.

In the Spectrum™ method, a universal intensity scale is used (Munoz and Civille 1992; Meilgaard and others 1999). Panelists are screened for cognitive, descriptive and sensory discrimination ability, interest, and availability. The panel size is similar to QDA. By this method, panelists score intensities in the same manner across all attributes. The scales are anchored on either end and can be 15-cm line scales or, more commonly, 0 to 15 numerical scales with tenth subdivisions between, yielding 150 points. The scales and their intensities have already been established, and during training, panelists are given established examples to familiarize them with the use of the scale. Although the scale anchors end at 15, panelists can assign higher numbers if necessary. The panel leader, who should have previous experience with the Spectrum™ method, takes an active role in panel training, including scale usage and language development and application. The panel, with assistance of the panel leader, identifies the sensory language for the product, their order of appearance, and definitions for each term. While a universal intensity scale is standard for the Spectrum™ method, product specific scaling can also be applied. As with QDA, data is readily analyzed by statistical techniques. Results are normally graphically presented using histograms.

Development of Flavor Lexicons

Characteristics of flavor lexicons have been previously discussed by Civille and Lawless (1986) and by Lawless and Heymann (1998). One key characteristic of a good flavor lexicon is that it be discriminating and descriptive. Ideally, for a descriptive language to be discriminating (able to differentiate the products for which it was developed), the language should be developed from a broad representative sample set that exhibits all of the potential variability within the product (Meilgaard and others 1999). Defining what constitutes representative is often a challenge, particularly with foods or commodities which are diverse in flavor, source and treatment. Standard descriptive languages have been developed for commodities using representative sample sets (Janto and others 1998; Johnsen and Civille 1986; Johnsen 1986; Johnsen and others 1987; Heisserer and Chambers 1993). In an extreme example, Drake and others (2001b) collected 220 samples of Cheddar cheese varying in age, milk heat treatment, and geographical origin to identify a descriptive language for Cheddar cheese. The sample set was screened to 70 cheeses prior to language generation.

Another key aspect of a flavor lexicon is for the terms to be nonredundant. That is, that multiple terms are not used to describe the same flavor or, alternatively, 1 term representing or overlap-
Flavor lexicons . . .

tive to the universal scale intensity reference. An example would be the use of a 16% sucrose solution to demonstrate a 15 on the Spectrum universal scale. The solution demonstrates the intensity of the 15, and all tastes and flavors are rated accordingly. These references are external to the product category or the specific attribute. All attributes are rated using this same yardstick, so that attributes in a product are comparative across products, across attributes in 1 product, and over time and different panels. A product specific scale rates attributes within that specific product category. An intense flavor within that product category is assigned a high intensity. The upper end of the scale represents the highest intensities within the particular product category. In this manner, products within the category can be compared but cannot be compared to other products. In attribute specific scaling, each attribute is rated independently so that each attribute has its own intense quantitative reference. An example would be fruity and sulfur flavor in Cheddar cheese. Each attribute would have a high quantitative reference for that attribute within that product category. Fruity flavor is not an intense flavor in Cheddar cheese, while sulfur flavor is. The quantitative references for these 2 attributes on an attribute specific scale in Cheddar cheese would be quite different in intensity if compared to each other. With this scale, an attribute can be compared across products, but attributes cannot be compared within a product. Attribute specific scaling is closest to what untrained panelists and consumers do. It is also important to note that the 3 types of scaling probably represent the most to least amount of panelist training, with the universal scale requiring the most amount of training and attribute scaling, the least training.

A key benefit of a descriptive lexicon is its ability to relate to consumer acceptance/rejection and instrumental or physical measurements. A true “holy grail” lexicon would be one that was readily related to both consumer and instrumental measures. In reality, establishing even 1 relationship is a challenge, and often is not possible for some products. When key flavor terms for a product’s category are represented by a group of chemicals that appear at low levels in the product (pyrazines and mercaptans for roasted coffee, peanuts, meat, chocolate), it is difficult to determine the exact mixture and amount of the various chemicals that characterize the “roasted” character. Also, no conclusions can be drawn regarding the nature of an attribute and whether the consumer response might be positive or not. For instance, the descriptor baby burp/vomit/butyratic acid could all be used to accurately describe 1 particular flavor. Alternatively, baby burp/vomit could be used and butyric acid could be used as the reference. The danger would be in assuming any relation to consumer acceptance or rejection of this flavor in a particular product without extensive consumer testing. The use of consumer terminology in a descriptive language does not mean the language has direct application to consumer acceptability. The apparent “negative” connotation of vomit or butyric can be misleading, since butyric is an inherent and appropriate description for many cheeses and other fermented foods. Further, consumer preferences and liking differ widely with age, geographic location, ethnic background, and familiarity with the product. What may constitute a “bad” flavor to one market segment may represent a “good” or “desirable” flavor to others.

In some cases, the lexicon has direct application to instrumental analysis. In the case of flavor lexicons, the lexicon has direct application to the volatile or nonvolatile compounds found in a product. Certainly, the use of chemical components, particularly chemical components isolated from that product, can make a lexicon more clear and grounded and can establish a link to formulation and/or production of that product. However, establishing those links can be a challenge and is addressed later in this review. Certainly, a lexicon can be representative, discriminative, and precise without using consumer terminology or having chemical references.

Application of Specific Flavor Lexicons

Flavor lexicons are used widely to describe and discriminate among products within a category. They are widely used in industry for comparing and monitoring products and product consistency and for profiling new and competitive products. The use of lexicons in quality control, with relationships to instrumental or consumer responses, is full of fruitful applications. Lexicons are also a powerful research tool with numerous applications across many commodities and commercial products. Descriptive analysis has been conducted widely on wine. Noble and others (1984) first described a standardized and referenced language for wine aroma. The language has been widely applied, expanded, and built on and used throughout the wine industry for research (Heymann and Noble 1987; Noble and Shannon 1987; Noble and others 1987; McDaniel and others 1987; McCloskey and others 1996; Siversten and Risvik 1994).

In commodities such as cheese, several different flavor lexicons have been used to document aroma and flavor development, the effects of fat reduction, and the effects of different starter or adjunct bacteria. Muir and others (1995a) described 9 aroma terms for characterization of aroma profiles of hard and semi-hard cheeses. Piggot and Mowat (1991) determined 23 descriptive flavor and aroma terms in a study of maturation of Cheddar cheese. McEwan and others (1989), Muir and Hunter (1992a), Muir and others (1996b) and Roberts and Vickers (1994) also developed descriptive flavor and aroma terms to study flavor during Cheddar cheese aging. Muir and others (1996a) and Drake and others (1996, 1997) used descriptive sensory panels to determine the effect of starter culture and adjunct cultures on Cheddar cheese flavor. Banks and others (1993) used descriptive analysis to determine sensory properties of low fat Cheddar cheese. A flavor lexicon for Comte cheese was used to identify naturally existing cheese georegions within France (Monnet and others 2000).

Sivertsen and Risvik (1994) demonstrated that French red wines could be differentiated according to region of origin using a sensory lexicon for wine. An aroma lexicon for meat/brothy aromas was used to optimize a meat-like process flavor made from enzyme hydrolyzed vegetable protein (Wu and others 2000). Several groups have identified and used descriptive sensory analysis to differentiate fluid milks (Lawless and Claassen 1993; Phillips and others 1995; Phillips and Barbano 1997; Watson and McEwan 1995; Chapman and others 2001; Bom Frost and others 2001). As might be expected, the languages have significant overlap in many terms. Flavor and aroma, as well as texture and feeling factors (aftertaste/mouthfeel) have been described. These lexicons have been used to determine the effects of fat content, storage, and other processing conditions on milk flavor and aroma perception. A flavor lexicon for chocolate ice cream was used to discern the effects of milkfat, cocoa butter, and fat replacers on sensory properties of chocolate ice creams (Prindiville and others 1999, 2000).

Descriptive analysis lexicons have also been used for numerous other products including almonds (Guerrero and others 1997), ripening anchovies (Bestiero and Rodriguez 2000), olive oil (Albi and Gutierrez 1991; Guerrero and others 2001) fermented milks (Hunter and Muir 1993; Muir and others 1997b), spoiled milk aroma (Hayes and others 2002), yogurt (Drake and others 2001a), beer (Meilgaard and others 1979), distilled beverages (Piggott and others 1980; Cristovam and others 2000; Lee and others 2001; McDonnell and others 2001) chicken (Lyon 1987), beef (Berry and others 1980), meat WOF (Johnsen and Civille 1986; Byrne and others 1999, 2001), Spanish cheeses (Ordóñez and others 2001).
As previously mentioned, the use of definitions and references is important in establishing a language that can be reproduced at different sites or at different points in time. Previous research on descriptive sensory analysis of Cheddar cheese (Piggott and Muir 1992), ham (Flores and others 1997), apples (Dalliant-Spinnler and others 1996), processed cheese (Muir and others 1997), sourdough bread (Lotong and others 2000), and chocolate (Lotong and others 2002) demonstrated that descriptive lexicons can be developed at different sites or at different points in time. These lexicons were used to document the development of different flavors in prototypes resulting from different ingredients, processes, and storage. At this phase, the lexicon often changes slightly to incorporate new attributes that develop in the new flavors. Once the final prototypes are ready for outside testing, the descriptive panel documents the products and prototypes going to the field test. Data relationships are established to determine the drivers of liking to set the best new product formulation and/or process. Once in production, these same attributes can be used to monitor production (Munoz and others 1992).

**Benefit of References and Definitions in Lexicon Application**

As previously mentioned, the use of definitions and references is important in establishing a language that can be reproduced at different sites or at different points in time. Previous research on descriptive sensory analysis of Cheddar cheese (Piggott and Mowat 1991; Roberts and Vickers 1994; Muir and Hunter 1992a; Muir and others 1995b) lacked definitions or references for descriptors. While some overlap between descriptors is observed, different research groups have used different descriptive vocabularies and rarely have definitions and references for descriptors been identified, making universal interpretation or reproduction of results difficult. A standard list of descriptors was identified for aged natural cheeses by Heisserer and Chambers (1993). More recently, Drake and others (2001b) identified a language for Cheddar cheese with standard definitions and references. Work is underway to confirm chemical references for this Cheddar cheese flavor lexicon. Murray and Delahunty (2000a) also developed and identified a referenced sensory lexicon for Cheddar-type cheeses. Both languages were developed independently, one in the U.S.A. (Drake and others 2001b) and one in Ireland (Murray and Delahunty 2000a). Studies have also been conducted to identify sensory attributes and descriptors for Parmesan and Comte cheeses (Virgili and others 1994; Berodier and others 1997).

Lexicons accurately communicate among research groups, particularly with a standardized lexicon and highly trained panelists. Previous studies with multiple panels at different sites have not used a uniform language. Risvik and others (1992) compared evaluation of chocolate between British and Norwegian panels, and Hirst and others (1994) later compared evaluations of cheese between British and Norwegian panels. Cross-cultural differences were attributed to the observed discrepancies in term usage and sample differentiation. Sensory evaluation of hard cheese has been continued in the European Union. Ring trials at 7 different sites across the EU were conducted, and a core sensory language for evaluation has been developed (Hunter and McEwan 1998; Nielsen and Zannoni 1998). However, definitions and references (anchors) for each descriptor were not identified. Lundgren and others (1986) conducted an interlaboratory international study on firmness, aroma intensity, sweetness, sourness, and flavor intensity of pectin gels using untrained panelists, and noted good agreement of the results, indicating that simple terms in simple systems may be universal in perception. Lotong and others (2001) reported the results of descriptive analysis of orange juice flavor by 2 highly trained descriptive panels using different descriptive analysis methods and independent lexicons. Similar patterns of differentiation among samples were observed between the 2 groups. Pages and Husson (2001) also studied panel performance with 13 different groups for 2 foods: compotes and chocolate. Efforts were made to standardize the descriptors and scale among the groups prior to the study, although the amount of panel training for each group was not specified. Again, similar patterns of differentiation were observed between the panels, although there were minor differences in scale usage and descriptor significance.

While similar patterns of differentiation among samples by panels that use different languages are expected (particularly if the lexicons are comprehensive and the panelists highly trained), standardized language with definitions and references improve communication, cross panel validation, and subsequent application of descriptive analysis results to instrumental or consumer data. Further, other sources of variation potentially exist in comparing panel results at different sites within the same country using the same language. Drake and others (2002) reported on the performance of 3 descriptive panels trained at different sites by different panel leaders on a previously developed and standardized cheese lexicon (Drake and others 2001b). Panels were able to accurately communicate attribute differences between cheeses. There were observed differences between these panels in scale usage and attribute recognition. These differences were attributed to differences in panel leadership and h of panelist training. In a similar study, Martin and others (2000) compared odor profile results of 2 panels. Language, scale, and method of presentation were standardized. Results obtained from the 2 panels were similar. Differences between attribute intensities were reported and were attributed to differences in individual panelist experience and/or perception. As with the conclusions of Drake and others (2002), strong panel leader interaction was recommended as a method to rectify these differences, along with regular feedback between the 2 panels.

**Relating Sensory Perception to Consumers**

While descriptive sensory analysis is used to identify and quantify information on the sensory aspects of products, effective consumer tests are used to provide information on consumer liking (Meilgaard and others 1999). Acceptability and preference tests provide quantitative consumer liking and/or preference information. Screeners and questionnaires are used to gather demographic data, frequency of usage, and purchase history about a particular product or group of products (Lawless and Heymann 1998; Meilgaard and others 1999). Questionnaires are often included with and are recommended with acceptance testing to aid in data interpretation. Despite collection of demographic and usage information, identifying drivers of liking or disliking within consumer market segments is critical for industry to identify which products and product attributes are preferred and by which consumer market segments.

In order to relate consumer liking and descriptive sensory properties, several statistical methods are applied. Simple techniques
Flavor lexicons . . .

such as scatter plots and correlation analysis, are critical to learning about the possible direct relationships between sensory attributes and consumer liking. Often, multivariate analysis is required to fully probe and relate the data sets and determine the combinations of attributes that affect liking. Multivariate techniques used include multiple linear regression analysis, canonical correlation analysis, principal component analysis, partial least squares regression analysis, and preference mapping, a principal component analysis derivation (Munoz and Chambers 1993; Heyd and Danzart 1998; Barcanes and others 2001). A group of products are documented with descriptive analysis and then tested with consumers. The analytical descriptive properties, which describe exactly what attributes are perceived and at what levels, are related to the consumer preferences using combinations of multivariate techniques. “Preference mapping”, as this relationship building is called, has been conducted with many products including cheese (McEwan and others 1989, Lawlor and Delahunty 2000; Murray and Delahunty 2000b; Barcanes and others 2001), chicken nuggets (Arthit 1997), coffee (Heyd and Danzart 1998), apples (Dallant-Spinnler and others 1996; Jaeger and others 1998), fermented milks (Muir and Hunter 1992b), sheepmeat (Prescott and others 2001), pudding (Elmore and others 1999), milks of varying fat content (Richardson-Harman and others 2000), and cured meats (Munoz and others 1995).

One of the biggest challenges in consumer research is the clarification of consumer language. Consumers may use terms that are ambiguous, have multiple meanings, are associated with “good” or “bad”, or are combinations of several terms. These integrated terms, such as “creamy”, are often used by consumers to represent a combination of positive attributes. Determining exactly what attribute or attributes “creamy” refers to (flavor compared with texture compared with mouthfeel) has been the subject of many studies relating consumer and trained sensory panels to clarify this term (Mela 1988; Elmore and others 1999; Bom Frost and others 2001). Dacremont and Vickers (1994a) used concept matching to clarify consumer perception of Cheddar cheese flavor. Descriptive analysis was not conducted to correlate with the obtained consumer information, and a limited number of consumers (n = 18) were used for the assessments. The concept of Cheddar cheese flavor is a consumer concept and likely varies considerably among consumers, as does Cheddar cheese flavor itself. A subsequent study was conducted by the authors using combinations of chemical volatiles (isolated from cheese) and concept matching to further clarify the role of individual and combinations of chemicals on the Cheddar cheese flavor concept (Dacremont and Vickers 1994b). Again, a limited number of subjects were used in the flavor concept experiments (n = 16 to 21), which potentially limits conclusions from the study. Doubtless, with large experiments and multiple combinations of treatments, the use of large numbers of consumers is impractical, but care must be taken when interpreting such data. Further studies with larger consumer groups and demographic information including types (brand, age) of Cheddar cheese normally consumed would provide additional clarification.

Relating Sensory Perception to Chemical Components

Relating sensory language and chemical volatile compounds represents a challenge for several reasons. Relative amount of a compound in a food is not necessarily a measure of its sensory impact, due to different thresholds and the effects of the food matrix. The sensitivity of the extraction technique must also be taken into account. Finally, only a small percentage of the volatile components in a food are odor-active (Friedrich and Acree 1998). The 1st step in the solution to relate gas chromatographic data and chemical compounds to sensory impact is gas-chromatography olfactometry (GC/O). In this technique, the individual compounds in the GC effluent are sniffed and described by a trained panelist. The technique can also be quantitative if the relative intensity of the odorant may also be recorded. Three different techniques are commonly applied with GC/O: aroma extract dilution analysis (AEDA), CharmAnalysis™, and OSME. These techniques and others have been loosely grouped into 3 categories: dilution analysis, posterior intensity methods, and detection frequency methods (van Ruth and O’Connor 2001). Both AEDA and CharmAnalysis operate on dilution of samples until an odor is no longer detected. The highest dilution at which the odor is detected can be converted to a flavor dilution value (FD value used with AEDA) or a Charm value (Friedrich and Acree 1998). The requisite sniffing of large numbers of sample dilutions by at least 2 panelists makes these techniques time-consuming. These techniques also assume that the response to a stimulus is linear and that all compounds have the same linear response. Osme, in contrast, is a time intensity technique and does not involve dilutions, but instead requires panelists (3 or more) to evaluate the time-intensity of aromas as well as the aroma character (Miranda-Lopez and others 1992; Plotto and others 1998; Fu and others 2002). Posterior intensity techniques can also involve 2 or more trained panelists simply scoring the maximum perceived intensity of an aroma along with its aroma character. A recent olfactometry technique, nasal impact frequency (NIF) has also been developed based on frequency of detection of a compound in the GC effluent by sensory panelists (Pollien and others 1997; Bezman and others 2001). The NIF technique can be applied with time intensity to yield surface of nasal impact frequency (SNIF) data. Pollien and others (1999) demonstrated that NIF and SNIF could be applied to quantify 1-octen-3-one in solution and in coffee aroma with results comparable to mass spectrometry. Both Osme and other posterior intensity methods and NIF/SNIF require fewer GC runs than AEDA and CharmAnalysis, since only the undiluted flavor extract is evaluated by the sensory panelist. However, a larger number of panelists (3 or more, compared to 2 for AEDA and CharmAnalysis) is recommended for these techniques. All of these techniques are useful for determining odor activity of volatile compounds.

One of the limitations to these techniques is using few panelists due to the time requirement. Usually a minimum of 2 sniffers is advised. Finally, the information obtained is based on individual components and does not include matrix effects or the time release factor involved when a sample is chewed in the mouth. Another device, the artificial mouth or RAS (retronasal aroma simulation) addresses the conditions encountered by food in the mouth (Roberts and Acree 1996; Friedrich and Acree 1998). Saliva, pH, temperature, and mixing can all be addressed, followed by subsequent analysis of the released volatiles in the headspace. Van Ruth and others (2001) studied the effects of saliva components of volatile release of 20 aroma compounds in model systems by static headspace analysis. Their goal was to relate results to volatile flavor release in the mouth. Their findings suggested that, while saliva composition does influence volatile release, the ratio of saliva to matrix played a larger role. Flavor release and perception is a very complex process (Taylor 1996). RAS simplifies a very complex process, and, while valuable insight is gained with this technique, the time release aspect of food consumption and subsequent flavor detection and the matrix effects are still not addressed.

Numerous studies beyond the scope of this review have been conducted on aroma volatiles of selected foods with GC and GC/O techniques. For the reasons mentioned previously, these results are useful but limited in any application to sensory perception (Piggott 1990). A list of volatile components with their odor properties does not equal sensory flavor. Descriptive analysis of the food product(s) can be done simultaneously with the instrumental
analysis and the results analyzed for statistical relationships by univariate and multivariate analysis of variance. Karagül-Yuceer and others (2001, 2002) conducted instrumental analysis (including GC/O) and descriptive sensory analysis of nonfat dried milk powders to determine potential relationships between instrumental and sensory perceptions of flavor. Cadwallader and Howard (1998) conducted a similar study of light activated milk flavor using instrumental analyses and descriptive sensory analysis, as did Young and others (1996) with apple flavor. And Mazetu and others (2001) contributed further flavor analysis. When a core taste, aroma, or flavor compound is incorporated into a food product, the resulting sensory perception is difficult to identify except for the flavor of interest is available, a side-by-side comparison can also be conducted. Suriyaphan and others (1999) identified alkadienals responsible for off-flavors in cheeses containing lecithin by a side-by-side instrumental comparison of experimental cheeses made with and without added lecithin. While potential relationships can be identified in this manner, further work is required to establish the significance of these findings. The subsequent steps in relating instrumental data to sensory perceptions of flavor are described in detail by Karagül-Yuceer, Acree, and others (2001a, 2001b; Kirchhoff and Scheiberle 2001). Compounds can be removed from the model mixtures sequentially and the similarity to the particular food or a model system followed by descriptive analysis provides further links to the actual role of the compound in human perception of flavor. Belford and others (1991) used descriptive analysis to determine the role of key volatile components of maple syrup, beef juice, and other foods. In this manner, contributing odorants were identified. Similar studies using model systems have been applied to strawberry juice, Gala apple flavor, and meat-like process flavoring, virgin olive oil, orange juice, and sourdough rye bread (Scheiberle and Hofman 1997; Plotto and others 1998; Wu and others 2000; Reiners and Grosh 1998; Buchet and Scheiberle 2001a, 2001b; Kirchhoff and Scheiberle 2001). Compounds can also contribute to flavor at concentrations at or below sensory thresholds, as was observed by interactions with other chemicals or the food matrix. Unfortunately, the role of these compounds is not addressed with this technique.

Panelists tasted water-soluble extracts of come cheese to identify fractions which had particular tastes in an attempt to clarify the role of peptides and amino acids on flavor (Salles and others 1995). Preininger and others (1996) used an unprinted cheese matrix to evaluate both volatiles and nonvolatile flavor components of 2 Swiss cheese samples. A similar study was conducted with Emmental cheese (Rychlik and others 1997). Suriyaphan and others (1999) identified key chemical volatile components of cow/berry and earth/bell pepper sensory perceived flavors in selected aged British Farmhouse cheeses. In this study, sensory properties were identified by descriptive sensory analysis, aroma volatiles were quantified by GC/MS, and aroma properties described by GC/O. Suspected key volatiles were selected from GC/O data based on aroma properties and FD values. The selected aroma components were subsequently incorporated into mild (bland) cheese across the concentration range encountered in the Farmhouse cheeses and evaluated by descriptive analysis. Studies such as these provide convincing evidence of the role of particular compounds on flavor. Model systems have not as yet provided insights into the role of compound mixtures and the role of compounds at sub-threshold.

## Conclusion

Flavor lexicons are an indispensable tool for accurately documenting the description of food flavor. As with any scientific method, sensory scientists need to adhere to those recommended practices required to assure panel performance: careful screening and selection of panelists; comprehensive training that includes lexicon development; definitive references; scaling standardization; experimental controls for sample preparation and evaluation; and ongoing maintenance of the panel system. Future research will continue to address relating descriptive (and analytical) lexicons with consumer acceptance and instrumental flavor analysis. More testing is expected and needed to demonstrate the robustness of the methodology and its practical application across cultures, languages, descriptive methods, and analytical chemical techniques.

## References


Flavor lexicons...
ZS 20010013. Submitted 1/8/02, Accepted 4/9/02, Received 4/10/02

Author Drake gratefully acknowledges the support of Dairy Management, Inc., California Dairy Re- search Foundation, and North Carolina State Univ. This manuscript is paper no. 02-52 in the journal series of the Dept. of Food Science, North Carolina State Univ., Raleigh, N.C. The use of trade names in this publication does not imply endorsement by the North Carolina Agricultural Research Service nor criti- cism of similar ones not mentioned.

Author Drake is with the Dept. Food Science, North Carolina State Univ., Raleigh, N.C., U.S.A. Author Civille is with Sensory Spectrum, Inc., Chatham, N.J., U.S.A. Direct inquiries to author Drake (E-mail: mdrake@unity.ncsu.edu).