

Towards a Taxonomy of Personality Descriptors in German: A Psycho-Lexical Study

ALOIS ANGLEITNER and FRITZ OSTENDORF
University of Bielefeld, Federal Republic of Germany

and

OLIVER P. JOHN
University of California at Berkeley, USA

Abstract

We present two studies aimed at developing a comprehensive taxonomy of German personality-descriptive terms. In the first study, all personality-descriptive adjectives (e.g. cynical), type nouns (e.g. cynic), and attribute nouns (e.g. cynicism) were extracted from a German dictionary. We found that almost half of all German adjectives were potentially personality-relevant, as contrasted with only 8% of the nouns. Moreover, there were more attribute nouns than type nouns, the latter appearing more slangy, metaphorical, concrete, and rich in imagery (e.g. Big-mouth, Wooden-head). In the second study, we discuss basic conceptual distinctions among units of personality description, develop a category system based on a prototype conception, and present a classification of 5092 adjectives into 13 categories. The classifications were generalizable across both judges and a two-year time interval, and agreed with a priori expert classifications. An analysis of the prototypical category cores suggested that Evaluations, Temperament and character traits, and Experiential states were represented most extensively in German, whereas Social effects, Roles and relationships, and Appearance were rather infrequent. These findings, though generally similar, differ from Norman's (1967) American taxonomy in the number of Evaluative terms and of Activity descriptors. Our studies provide comprehensive and representative lists of German words for personality traits, moods and emotions, social roles, effects, evaluations, and physical appearance, and may serve as the basis for taxonomies, dimensional analyses, and assessment instruments. We emphasize the need to standardize procedures in taxonomic research and outline suggestions for future studies of other languages.

INTRODUCTION

What exactly do we mean by *personality*? There are probably as many answers to this question as there are respondents. Most personality textbooks 'compare and

contrast' dispositional, behaviorist, dynamic or motivational, cognitive, role-theoretical, and many other 'approaches' to the study of personality. The wide range of possible units of personality description is also apparent in the definition of personality 'in its widest reference as referring to individual differences in temperament, intelligence, character, attitudes, aptitudes, etc,' a definition taken from the aims and scope of the journal *Personality and Individual Differences*. In the present paper, we will be concerned with all these units of personality description, and we will pay some particular attention to the units so conveniently summarized by 'etc.' in the definition above.

More specifically, we are interested in documenting and systematizing what it is that can be said about persons and their personalities in one particular language. The personality descriptors in a language can be sampled in many different ways. For example, one may obtain unstructured personality descriptions from a sample of subjects (e.g. Bromley, 1977; Fiske and Cox, 1979; see John, 1990), or one may ask the subjects to recall conversations about their own and others' personalities (e.g. De Raad, 1984). Whereas these procedures have the advantage of identifying concepts that the people in that sample have actually used, the representativeness of these concepts for other samples and under different eliciting conditions is difficult to establish. Indeed, the particular concepts subjects will use at any given time vary widely as a function of both the target being described (John, Hampson and Goldberg, 1990), and the characteristics and goals of the describer (e.g. Fiske and Cox, 1979). That is, in any particular sample and at any particular time, only a small fraction of the concepts that *can* be used will actually be used to describe personality.

THE LEXICAL APPROACH

An extensive, if not exhaustive, sample of what can be said about personality in a language is contained between the covers of an unabridged dictionary, providing a sample of terms compiled over many years and updated and refined by generations of lexicographers.¹ In the present project, we have followed a research tradition we refer to as the *lexical approach* and have used a dictionary to obtain a first comprehensive listing of personality descriptors.

In a recent historical review, we have described the procedures and findings emanating from lexical research on personality descriptors in American English and in Dutch (John, Angleitner and Ostendorf, 1988). In general, researchers in the lexical tradition have begun their research with an analysis of the dictionary, initially extracting an exhaustive set of personality-descriptive terms and subsequently reducing it according to a number of criteria. Most studies have focused on personality traits, rather than states, activities, or social effects, and have sought, typically via factor analysis, a few broad dimensions to represent the total set of descriptors.

¹ The individual concepts that constitute the contents of the personality dictionary in a language obviously do not exhaust everything that can be said about personality in that language. Dictionaries provide the building blocks, the concepts, that can be combined into sentences in nearly infinite ways to capture something new, unique, or subtle about an individual. However, a better understanding of the nature of the individual concepts is needed before we can study, as McCrae (1990) suggests, person descriptions that are even more complex, such as sentences and paragraphs.

THE BIG FIVE DIMENSIONS IN STUDIES OF DIFFERENT LANGUAGES

At the broadest level of abstraction, a five-factor structure—often referred to as the ‘Big Five’ (Goldberg, 1981)—seems to provide a replicable representation of the major dimensions of personality-trait descriptors in English. When large and representative sets of variables are factored, the same five factors (Surgency or Extraversion, Agreeableness, Conscientiousness, Emotional Stability versus Neuroticism, and Intellect/Openness) seem to emerge reliably, across different types of samples, raters, and methodological variations (for reviews, see John, 1990; McCrae, 1989).

The generalizability of this dimensional structure has been explored in some initial studies using languages other than English. Such research can determine the applicability of the five-factor taxonomy in other cultural contexts and test the hypothesis that there are universals in the encoding of individual differences across languages and cultures (e.g. Goldberg, 1981). The existence of cultural universals would be consistent with an evolutionary interpretation of the way in which individual differences have become encoded into the natural language: if the tasks most central to human survival are universal, the most important individual differences, and the terms people use to label these individual differences, would be universal as well (Borkenau, 1990; Hogan, 1983).

Although central from the vantage point of the lexical approach, cross-language research is difficult and expensive to conduct, and consequently rare. In most comprehensive taxonomic studies, English has been the language of choice, primarily because the taxonomers were American. The only non-English taxonomy project that began with an analysis of the dictionary is the Dutch project, carried out by Hofstee and his colleagues at the University of Groningen in the Netherlands (Brokken, 1978; De Raad, Mulder, Kloosterman and Hofstee, 1988; see also John *et al.*, 1988, for a review). Extensions into cultures different from the industrialized West are just beginning to appear. White (1980) studied the structure of interpersonal traits in the A’ara (Solomon Islands) and Orissa (India) languages, and Bond and his collaborators (Bond, 1979, 1983; Bond and Forgas, 1984; Bond, Nakazato and Shiraishi, 1975; Nakazato, Bond and Shiraishi, 1976) compared the personality factor structures of Hong Kong and Japanese samples with previous analyses of U.S. and Filipino (Guthrie and Bennett, 1971) data.

A recent and more extensive study of Filipino samples provided further support for the lexical universality hypothesis. Church and Katigbak (1989) used an emic (culture-specific) strategy in sampling descriptors in both languages, rather than simply translating English descriptors into the Filipino language under study. This study provides the best evidence so far for the comprehensiveness and relevance of the Big Five dimensions in a non-Western culture. As the authors caution, however, ‘this does not mean that there are no unique concepts in either language. However, at a higher level of generality, similar structural dimensions emerge’ (Church and Katigbak, 1989, p. 868).

PREVIOUS RESEARCH IN GERMAN

More than 50 years ago, Baumgarten (1933) published the first extensive listing of personality descriptors. Whereas her research stimulated no subsequent research

in German, it influenced Allport and Odbert's (1936) classic analysis of personality descriptors: *Trait names: A psycho-lexical study*. Subsequent work on English and Dutch descriptors eventually rekindled interest in German terms, and in 1984 we reported an initial study of German-American bilinguals (described in John, Goldberg and Angleitner, 1984). A set of 162 trait adjectives was selected to be representative of the American trait taxonomies, was translated into German, and used to obtain self-descriptions from a sample of 70 American-German bilinguals. The unique advantage of the bilingual design, in which the same subject provides descriptions in both languages, is that sample differences can be controlled and that translation checks can be made at the level of individual items. Indeed, we found that, although translation adequacy was generally good, there were a few English terms that did not translate well into German (John *et al.*, 1984). Because the size of the bilingual sample did not permit factor analyses, we evaluated the cross-language generalizability of the Big Five in terms of (a) the internal consistency of the hypothesized German markers of each of the five dimensions, and (b) a multi-trait-multilanguage matrix. The internal consistency (Coefficient Alpha) of the Big Five scales scores, each based on 16 rationally selected adjectives, were similar in the two languages, with a median of 0.82 for the English and 0.77 for the German scales. More important, the cross-language consistency correlations (i.e. monotrait-heterolanguage coefficients) ranged from 0.72 to 0.84.

This research, although consistent with the hypothesis of universals in the language of personality, was limited in several ways. For one, the number of descriptors that could be studied in a bilingual design was necessarily small. More important, however, given the lack of any systematic research on the dimensions of personality in German, we could not follow the emic (culture-specific) research strategy used by Church and Katigbak (1989). Instead, the 162 trait adjectives and the Big Five concepts were 'imported' from one language (English) into the other (German), hoping that these 'imported' concepts would suffice to capture the range of personality-descriptive meaning in German. Indeed, as Yang and Bond point out, in most cross-language studies,

'raters outside America have been invited to use the possibly Procrustean bed of personality descriptors taken from the English language and then processed by Americans. It has not yet been determined how dimensions derived by this use of the American materials might overlap with those derived from that culture's own language system' (1989, p. 3).

In other words, an understanding of the dimensions of personality description in German must begin with an analysis of German terms and generate a language-specific taxonomy, which can then be compared to the findings from the American and Dutch analyses. The present paper reports the first two studies towards that goal.

STUDY 1: ANALYSIS OF THE GERMAN DICTIONARY

Issues in the construction of taxonomies

The major issues that have to be addressed in the construction of a personality taxonomy have been discussed by John *et al.* (1988). First, the taxonomers must

specify the domain of phenomena to be covered by the taxonomy. This task is inherently theoretical, tantamount to defining what one means by personality. Taxonomers have to make practical decisions concerning their definition of personality—they have to specify what should be classified before they can begin classifying it. After the domain to be included in the taxonomy has been specified, one still has to assemble the universe of instances that fit that specification and decide how to select individual instances from that universe. That is, in the second step the taxonomers have to develop operational rules that can serve to identify instances considered relevant to the domain specification. These concerns are addressed in our first study. We define the universe of instances as all those adjectives and nouns in the German dictionary that are relevant for personality description. We also develop some initial operational criteria of ‘personality relevance,’ which we further specify and explicate in our second study, examining 13 categories of personality description.

In addition to domain specification and instance identification, the scope, or inclusiveness, of the classification has to be determined. Should one seek a comprehensive taxonomic structure that covers all types of personality characteristics, or would a taxonomy covering only a restricted domain be more useful? We have considered several different domains of personality description within one overall scheme but plan to subsequently examine the structure of each domain separately. A final set of issues involves the evaluation of the taxonomy at each stage. We will be primarily concerned with the comprehensiveness of the representation of concepts and with the generalizability of judgmental data across judges, time, and an expert criterion.

Three classes of personality terms: adjectives, type nouns, and attribute nouns

Most previous research has focused on personality-descriptive adjectives. In the two major compilations of personality descriptors, Allport and Odbert (1936) and Norman (1967) generally preferred adjectives and included nouns only when there was no corresponding adjectival form in the dictionary. The only extensive collections of nouns have been constructed by Goldberg (1981, 1982; but see also De Raad and Hoskens, 1990) who informally collected about 2,000 English personality nouns, such as *bloodsucker*, *nitwit*, and *jerk*. Each of these nouns refers to a set of persons who have in common either a particular trait or a constellation of traits. These nouns refer to personality *types* and, as Cantor and Mischel (1979) have shown, share some properties with the object categories (e.g. bird, tree) studied by Rosch (1978). Although hierarchical relations among a small number of type nouns have been examined (Cantor and Mischel, 1979), their content and dimensionality have not been studied systematically.

Type nouns are not the only personality-relevant nouns that appear in natural languages. John *et al.* (1984, p. 85) distinguished type nouns from a second class of nouns denoting the *attributes* of individuals. Attributes are used frequently in everyday life: We compare individuals with respect to their *friendliness*; we value *honesty*; and we abhor *violence*. These terms are abstract nouns that do not refer to people but to their traits, behaviors, and experiences. Indeed, they are the kind of variables most commonly studied in the personality literature; extraversion, intelligence, and private self-consciousness are all attribute nouns. In the present project, we retained these attribute nouns, so as to make them available for subsequent and more detailed study. In all, then, we considered separately three word classes,

namely adjectives (e.g. cynical), type nouns (e.g. cynic), and attribute nouns (e.g. cynicism).

Defining personality relevance

As a general definition, we considered *personality relevant* all those terms that can be used 'to distinguish the behavior of one human being from that of another' (Allport and Odbert, 1936, p. 24). To permit a group of judges to apply this rather abstract definition to a set of candidate terms, we elaborated this definition in three different ways. First, we identified, on the basis of the American and Dutch taxonomies, six kinds of person characteristics we expected to find in the German dictionary, including stable traits (e.g. friendly), states and moods (e.g. irritated), activities (e.g. hesitating), social roles, relations, and effects (e.g. brotherly), abilities and talents (e.g. capable), and appearance and physical characteristics (e.g. tall).

Second, we specified several exclusion criteria. The first exclusion criterion was the inverse of the general definition, specifying that a term is *not* personality relevant if it is nondistinctive and applies to all individuals (e.g. born, breathing, human). Also excluded were terms referring to geographic origin (e.g. Prussian), to nationality (e.g. Spanish), and to professional or job-related identities (e.g. physician, student), as well as terms that refer only to a part of the person (e.g. shining eyes) or terms whose personality implications are both metaphorical and tenuous (e.g. mouse, rose).

A third way of elaborating the definition of personality relevance was suggested by the Dutch taxonomy team (Brokken, 1978). To provide the judges with a linguistic test for deciding whether a term is personality relevant, we developed several heuristic criterion sentences. Adjectives that are personality relevant should fit in either of the following two sentences (some examples are given in brackets):

- (1) How [adjective] am I? [upset, attractive]
- (2) How [adjective] did Chris behave? [outgoing]

For type nouns, we used the two sentences:

- (3) Is Chris a [noun]? [nerd]
- (4) Can you call him/her a [noun]? [chicken]

Finally, attribute nouns can be identified by either of these two sentences:

- (5) Chris's [noun] is remarkable. [friendliness]
- (6) Does he/she have or possess [noun]? [patience]

Considering each term in these six sentence contexts helped the judge (a) identify the three relevant word classes (i.e. adjectives, type nouns, and attribute nouns), and (b) decide whether a term can be used for personality description at all. In general, we used a lenient criterion in our operational definition of personality relevance. During this first stage of the project, our general goal was to retain all German words that could possibly be used for personality description. We thus instructed our judges to be overinclusive and to retain a word even when they had doubts about its personality relevance.

Method

Subjects

Seven advanced graduate students in psychology and one Diplom psychologist served as judges. All 8 judges had extensive training in personality psychology but were not familiar with this particular project. None of the authors served as a judge.

Materials and procedure

We used the newest available edition of a comprehensive German dictionary (Wahrig, 1981), containing about 97,000 separate entries. To make the task less daunting for the judges, the dictionary was divided into 10 parts, each 80 to 100 pages long and rebound as a separate booklet. Each booklet was scanned for personality-relevant terms independently by two judges. Overall, the graduate student judges each worked through two different booklets, whereas the psychologist serving as a judge examined six different booklets.

The judges were given extensive instructions² defining the three syntactic forms of personality descriptors (i.e. adjectives, type nouns, and attribute nouns), and the three ways of elaborating the definition of *personality relevance*. As a general rule, however, the judges were told, 'If in doubt, always write the word down.' The judges extracted from the dictionary each word they considered personality relevant, keeping separate lists for the three word classes. As an initial check on the familiarity of the words, the judges also rated their familiarity with the word on a 3-step scale and indicated whether the word was marked in the dictionary as being 'out of use' (veraltet).

Results and discussion

A sample of pages from the *Wahrig* dictionary suggested that, as in dictionaries of English, the vast majority of lexical entries are nouns. As shown in Table 1, we estimated that the *Wahrig* dictionary contains more than 77,000 nouns (approximately 80% of the total corpus of German words) and about 12,000 adjectives (12%). The remaining 8,000 entries represent several other word classes, including verbs as well as pronouns, articles, and conjunctions.

Frequencies of potentially personality-relevant terms

Given our goal to be as inclusive as possible, we decided to retain as *potentially* personality-relevant every term that at least one judge had categorized that way. The frequencies of personality adjectives, type nouns, and attribute nouns are summarized in Table 1. Among the 11,600 German adjectives, 4,827 (42%) were categorized as personality relevant by at least one of the two judges who had examined that word; 2,658 terms were listed by both judges. The Coefficient Alpha reliability of the combined judgments was 0.73, a level quite acceptable given the difficulty and length of the task and the general instruction to err on the side of overinclusiveness.

² The complete instructions (in German) are available from the authors.

Table 1. Analysis of the German dictionary: Frequencies of three types of personality descriptors

Word class	Number in dictionary	Number possibly personality-relevant	Interjudge agreement (Coefficient Alpha)
Adjectives (e.g. Cynical)	11,600	4,827 (42%)	0.73
Nouns	77,330	5,819 (8%)	0.72
Types (e.g. Cynic)		2,212 (3%)	0.67
Attributes (e.g. Cynicism)		3,607 (5%)	0.76
Total nouns and adjectives	88,930	10,646 (12%)	0.76
All other word classes (e.g. verbs)	7,734		
Total: All words	96,664		

The number of personality-relevant adjectives in German appears to be considerably smaller than in either English (probably more than 15,000; see Norman, 1967) or Dutch (about 8,700; see Brokken, 1978). This difference is probably a function of the generally smaller size of the lexicalized vocabulary in German as compared with English, rather than an indication that German contains a smaller percentage of person-descriptive terms. Indeed, the finding that almost a half (42%) of all German adjectives is potentially personality-relevant might be surprising at first. However, analyses of the adjective lexicon in English seem to confirm this finding; about one-half of the English adjectives studied by Gross, Fischer and Miller (1989) can be applied to persons, their behaviour, experience, and appearance. More generally, personality seems to be one of a limited number of domains that are lexicalized as adjectives (if the language has any). Indeed, Dixon (1977) included 'human propensity' terms as one of the semantic types that languages universally express with adjectives.

For personality type nouns, the combined judgments had a Coefficient Alpha reliability of 0.67, slightly lower than the Alpha of 0.76 for attribute nouns. In all, 2,212 type nouns and 3,607 attribute nouns were judged personality-relevant by at least one judge. That is, only 8% of the German nouns were deemed potentially relevant for personality description, the class of attribute nouns (5%) being slightly larger than the class of types (3%). The finding that only 8% of the nouns but 42% of the adjectives were personality relevant is consistent with the idea that the German personality lexicon consists primarily of adjectives. Unfortunately, the American (Norman, 1967) and Dutch (Brokken, 1978) taxonomers did not report the frequencies of personality descriptors relative to the frequencies of the different word classes in the dictionary as a whole, and direct comparisons are therefore difficult. Nonetheless, our finding that there are relatively more personality-descriptive adjectives than there are type nouns seems to hold in English as well (Goldberg, 1982), despite the fact that in both English and German dictionaries nouns vastly outnumber adjectives.

When adjectives and both kinds of nouns are considered jointly, about 12% of

the total German vocabulary are potentially personality-relevant. This number contrasts with an estimate of 5% derived from Norman's (1967) study of American English. This apparent discrepancy is due to our systematic inclusion of both type and attribute nouns, whereas Norman (1967) generally preferred adjectival forms over nouns, retained type nouns only when a related adjective was not in the dictionary, and did not include attribute nouns at all. The closest approximation of the American estimate in our data is the number of adjectives relative to the total corpus of dictionary entries, a fraction that comes to 5%, exactly the same number as we estimated from Norman's (1967) research.

Comprehensiveness of the master list of German terms

To further examine the reliability and comprehensiveness of the lists extracted from the dictionary, we tried to identify adjectives possibly overlooked by our judges or omitted by systematic biases in the way terms are included in the German dictionary. The initial pool of 4,827 adjectives was checked for completeness against several lists of adjectives, including the first list of German personality-descriptors ever published (Baumgarten, 1933), a German translation of the Adjective Check List (Gough and Heilbrun, 1965), a German list of mood and emotion descriptors, called *Eigenschaftswörterliste* (Janke and Debus, 1978), an Adjective List for Self-Description (Mummendey, Mielke, Maus and Hesener, 1977), a set German trait-descriptive adjectives, the German Adjective List (John, 1982), and the terms included in psycholinguistic research by Wippich and Bredenkamp (1977).

Comparisons with the more than 2,200 terms on these lists identified a set of 265 potentially personality-relevant adjectives (i.e. an additional 5% of the total pool) that were either not listed in the *1981 Wahrig* dictionary or listed there only as nouns or verbs. Initially, we were surprised by the number of additional adjectives. However, an analysis of these words showed that they were unusual in several ways. First, several of the lists were research instruments that had been *translated* from English into German and thus included literal translations that do not exist as single words in German (e.g. 'selbstaufmerksam' as a translation of self-aware). Second, a substantial percentage of the additional terms were technical terms representing psychological jargon (e.g. *selbstdiszipliniert* [self-disciplined]) not (yet) part of the everyday language represented in the *1981 Wahrig* dictionary. Third, the remaining adjectives included (1) newly formed compounds, such as *auseinandersetzungsfreudig* (enjoys confrontations), (2) verb participles not listed separately in the dictionary but of considerable interest as potential person descriptors, such as *aufbrausend* (agitated), and (3) derivatives from nouns, such as *partnerschaftlich* (like a partner). Overall, these additional adjectives may be potentially useful for personality description although they are probably less common and familiar than those listed in the dictionary. We therefore added the additional 265 words to our Master List, resulting in a total of 5,092 adjectives available for subsequent research.

In conclusion, this comprehensiveness check suggests that our analysis of the dictionary identified more than 90% of all potentially person-descriptive adjectives in German. However, there are potentially personality relevant terms that will not be found in the dictionary. In German, these include psychological jargon that is generally understood, compound words, and adjectives formed from verbs (e.g. participles) and nouns already listed in the dictionary. These findings suggest that in

future research taxonomers should pay increased attention to differences between word classes (e.g. nouns versus adjectives), to the lexicalization rules of the language under study (e.g. in German, the participles of verbs are not considered separately as adjectives, and only a subset of all possible compound words can be listed in the dictionary), and to differences in the kinds of terms included in dictionaries and those included in other sources, such as assessment instruments constructed by psychologists.³

Hypotheses about type nouns

Although we found them particularly difficult to translate into English (or, for that matter, into any other language), the German type nouns seem to have characteristics similar to English type nouns (Goldberg, 1981, 1982; see also De Raad and Hoskens, 1990). Terms whose approximate English translations are Braggart, Devil, and Big-mouth (see Table 2 for the German terms) are more informal, more slangy, and (in German) more regionally (or dialect specific) in their use than are either attribute nouns or adjectives. The finding that 13% of the type nouns were unfamiliar to our judges, compared to only 6% of the attribute nouns, seems consistent with this characterization. The type nouns also seem much more concrete and rich in imagery (e.g. Big-mouth, Wooden-head) and often invoke metaphors. For example, the German type noun 'Waschlappen' brings to mind the image of a wet wash cloth and, when applied to a person, the lack of an inner backbone characteristic of a 'wimp'. These observations are consistent with Cantor and Mischel's (1979) finding that type nouns (e.g. comic joker) tend to elicit far more imagery and richer associations than do personality attributes (e.g. extraverted person). Empirical research is now needed to test the hypothesis suggested by these informal analyses. In particular, we suggest that type nouns are more extreme and polarized in their evaluation than are adjectives (Goldberg, 1982) and attribute nouns, are used more in oral discourse than in writing (John *et al.*, 1984), and serve to emphasize the speaker's emotional and evaluative stance towards the target person.

Familiarity of terms in the master list

We obtained ratings of familiarity from our judges, so that we would be able to exclude infrequently used and unfamiliar terms from further consideration. However, only 8% were categorized as unfamiliar by at least one judge, and only 2% of the personality-relevant terms were coded as 'out of use' in the *Wahrig* dictionary. These numbers are too low to significantly reduce the total pool of terms, and we therefore decided to retain all terms in the masterpool at this stage. We have constructed an alphabetical listing of the terms in each of the three word classes, including each term along with the number of judges rating it as personality relevant and with

³ Our analyses suggested that verbs have potential relevance to personality. Verbs are an integral part of person descriptions in personality questionnaire items, such as 'she likes to party' (see Angleitner, John and Löhr, 1986) and in descriptions of behavioral acts, such as 'she went to the party' (Buss and Craik, 1983). We did not include verbs (e.g. talk, like, party) as a separate word class because, in contrast to adjectives and nouns (e.g. talkative, likable, partier), verbs cannot be applied directly to individuals. For a discussion of the personality implications of verbs, see DeRaad *et al's* (1988) analysis of personality-relevant verbs in Dutch.

its familiarity rating. To illustrate the nature and composition of these lists, a sample of 20 terms from each word class is given in Table 2.

Table 2. Samples of terms from the masterpool of 4,827 personality descriptive adjectives, 2,212 type nouns, and 3,607 attribute nouns

Personality adjectives	Type nouns	Attribute nouns
abartig	Blagueur	Annehmlichkeit
aufsehenerregend	Deubel	Biederkeit
brummig	Erzieher	Einsamkeit
einfühlsam	Friedensschaffer	Fremdheit
feige	Grossmaul	Gesetztheit
geistreich	Holzkopf	Herzbeklemmung
grosstuerisch	Kanzelschwalbe	Intoleranz
homoerotisch	Kontaktmann	Konservatismus
klassisch	Lämmlein	Langmut
lautlos	Luder	Noblesse
misstrauisch	Naturkind	Randal
panisch	Pfeife	Schönheit
resch	Quacksalber	Stilgefühl
schöntuerisch	Sackerloter	Überdruss
speerig	Schlucker	Unmoral
traditionsbewusst	Sorgenkind	Unvorsichtigkeit
unehrlich	Techniker	Verbohrtheit
unrettbar	Trödel Fritz	Verworfenheit
unzuverlässig	Verdinger	Willenskraft
zerstreut	Waschlappen	Zwanglosigkeit

STUDY 2: CLASSIFYING ADJECTIVES INTO PERSON DESCRIPTIVE CATEGORIES

To be of practical value, a taxonomy of German personality descriptors must provide more than an alphabetical listing within three different word classes. A useful taxonomy must provide a systematic framework for distinguishing, ordering, and naming individual differences in people's behaviour and experience (John, 1989, 1990). Therefore, in our second study, we classified the terms into 13 categories of personality description, including traits, states, activities, social effects, and appearance. The basic assumption guiding this classification is that an individual's personality may be described with different conceptual units and at different levels of abstraction (Briggs, 1989; Hampson, John and Goldberg, 1986; John, 1989). As we will argue below, individuals can be described by their enduring *traits* (e.g. irascible), by the *internal states* they typically experience (furious), by the *physical states* they endure (trembling), by the *activities* they frequently engage in (screaming), by the *effects* they have on others (frightening), and by general *evaluations* of their conduct by society (unacceptable, bad). In addition, individuals differ in their anatomical and

morphological characteristics (e.g. short) and in the personal and societal evaluations attached to these appearance characteristics (e.g. pretty).

Categories suggested in taxonomic studies

The easiest way to illustrate the taxonomic distinctions initially proposed by Allport and Odbert (1936) and by Norman (1967) is to consider one particular personality characteristic across several categories. Let us take as an example the characteristic irascible, classified by Norman (1967) as a stable trait. At any given moment in time, this disposition may or may not be manifested in the individual's momentary condition. Norman (1967) differentiated two kinds of such conditions, states and activities. *Furious*, for example, is a state descriptor, whereas *screaming* is an activity descriptor. In the present study, we have added another category, physical and bodily states; *shaking* and *trembling*, for example, are physical states that may accompany both the experience of intense anger and behaviors such as screaming and yelling. Although the same trait may give rise to all three kinds of temporary conditions, they are distinct from each other and from the trait itself (Chaplin, John and Goldberg, 1988).

A third class of terms describes the effects that the expression of traits in behavior and emotion has on other people. Terms such as *frightening* and *intimidating* do not refer to traits but to the effects *irascible* individuals have on others; these terms signify an individual's 'social stimulus value' or reputation (Allport and Odbert, 1936). The mediating link between trait and social stimulus values are the individual's behaviors and emotional experiences; they constitute the stimulus that influences others in the individual's social environment. The natural language also includes descriptors that are even more evaluative than these social-effect terms—for example, *horrid*, *terrible*, and *bad*. Norman (1967) argued, as had Allport and Odbert (1936), that these terms contain too much evaluation and too little descriptive meaning to be useful for descriptive purposes, and thus assigned them to an exclusion category. In the present taxonomy, these terms have been retained as a separate category for subsequent study.

The distinctions reviewed so far are similar to those suggested by Allport and Odbert and by Norman. In addition, we have examined several other proposals for classifying the ways in which personality can be described, and in Tables 3 and 4 we present a detailed classification system integrating these earlier efforts. In Table 3, we summarize Norman's taxonomy and four of the other category systems we reviewed, and compare them to the categories we are proposing here. In Table 4, our categories are listed in more detail and illustrated with an example.

Wiggins (1979) was concerned, as was Norman (1967), with delineating distinctions among terms that can be used to describe differences among individuals. The six major classes outlined by Wiggins are generally similar to the ones discussed above: (1) Anatomical and physiognomic differences, (2) Personal appearance, grooming, and clothing, (3) Social background, roles, and demographics, (4) Effects on others (i.e. social stimulus value), (5) Temporary characteristics subdivided into current States, Moods, Attitudes, and Activities, and finally (6) Traits, which Wiggins differentiated into Interpersonal, Material, Temperament, Social roles and status, Character, and Mental abilities. As shown in Table 3, we have adopted Wiggins' proposal to further differentiate within the class of traits, although we have retained only

Table 3. Review of category systems to classify kinds of person characteristics

Categories proposed here	Norman (1967)	Wiggins (1979)	Guilford's Star Model	Questionnaire items ^a	Free Person Descriptions ^b
(1a) Temperament & character traits	Stable 'biophysical' traits (Cat. 1-3)	Traits—4 types (6, a, b, c, e)	Temperament	Trait attributions (2)	Properties (Traits & causes) (6a, b)
(1b) Abilities, talents, or their absence		Mental abilities (6 f)	Aptitude		
(2a) Emotions, moods, and cognitions	Temporary states (4)	States, moods (5 a, b)	Needs Interests	Covert reactions (1b) Wishes & interests (3)	Interests (6d)
(2b) Physical and bodily states			Physiological differences	Physical symptoms (1c)	
(2c) Behavioral states: Observable activities	Activities (5)	Activities (5 d)		Overt reactions (1a)	Behavior (2)
(3a) Roles and relationships	Social roles (8)	Roles & status (6d)		Biographical items (4)	Relationships (3)
(3b) Social effects: Reactions of others	Social effects (9)	Effects on others (4)			
(3c) Pure evaluations	Evaluative terms (12)				
(3d) Attitudes and worldviews		Attitudes (5c)	Attitudes	Attitudes (5)	Attitudes (6c)
(4a) Anatomy, constitution, morphology	Anatomical and physical terms (13)	Anatomy (1)	Morphological differences		
(4b) Appearance, looks, deportment		Appearance (2)			Appearance (1)
(5a) Context-specific or technical terms				Others (6)	
(5b) Metaphorical, vague, outmoded terms	Ambiguous Obscure (14, 15)			Bizarre (7)	
Additional content categories		Social background, roles, and demographics (3)			Origin (5) Context (4)

^aAngleitner, John and Löhr (1986, p. 69). ^bFiske and Cox (1979, pp. 143-144).

three of his six distinct categories. In particular, we have included in one category (1a) not only temperament and character but also interpersonal and material traits, whereas mental abilities (1b) and roles and relationships (3a) are treated separately.

Other classifications of personality descriptors

Whereas Wiggins' (1979) distinctions originated in the same taxonomic context as Norman's, Guilford (1959) postulated seven different personality domains, which he arranged in the form of a star; at the center of the star is the concept of personality. Although Guilford's seven domains correspond well with the categories discussed so far, his system led us to include in our category Experiential states (2a) not only moods and emotions but also cognitive and cognitive-affective states, such as interests and needs. Moreover, we concluded that attitudes, opinions, and worldviews (3d) should be considered a separate class of social or societal characteristics.

Whereas Guilford's distinctions were entirely theoretical, Angleitner, John and Löhr (1986, pp. 69–71) constructed their category system on the basis of a content analysis of items included in personality questionnaires. The vast majority of such items ask the respondent to report about various kinds of reactions. These reactions may be Overt (i.e. observable behaviors, such as habits and typical activities), Covert (i.e. private and not observable, such as internal sensations, feelings, and cognitions), or Physical (e.g. sweating, shaking). Other, less frequently found item types were Trait attributions, Wishes and interests, Biographical facts, Attitudes and beliefs, Others' reactions to the respondent, and 'Bizarre' items describing unusual, strange, or even abnormal behaviors (e.g. 'somebody is trying to poison me'). Despite the rather different purpose for which these categories were developed, each of them is represented in the system proposed here.

Studies of free person descriptions by children and adults have also led to several taxonomies of person descriptors (e.g. Bromley, 1977; Livesley and Bromley, 1973). In an empirical study, Fiske and Cox (1979) tested the validity of a taxonomy of terms used in open-ended person descriptions. Their taxonomy, which is included in Table 3, was rationally derived from a category system Anglin (1977) developed to account for children's descriptions of objects. The categories subjects used most frequently in Fiske and Cox's (1979) free description tasks were Appearance and 'Properties.' Fiske and Cox's Appearance category combined body, physique, and face—thus corresponding to our category (4a)—with grooming, clothing, and attractiveness—our category (4b). Similarly, as shown in the last column of Table 3, the category labeled Properties included several of our categories, in particular, Traits and causality (1a, 1b), Attitudes (3d), as well as Interests (2a). The category Relationships includes elements from our category (3a), such as role, social network, and history, and category (3b), such as the perceiver's reaction. Moreover, Fiske and Cox's category Behavior closely matches our category (2c). Two additional categories, Origin and Context, were almost never used either by Fiske and Cox's adult subjects or by children (Livesley and Bromley, 1973), and were therefore not included in the present study. Indeed, we had already excluded descriptors of national, ethnic, and regional origin in the first stage of this research. Other aspects of origin, such as social class, family, and educational background would be classified in our category Roles and relationships (3a).

Table 4 provides an overview of the final form of the category system as it was

applied in the classification of our Master List of 5,092 German adjectives. The 13 specific categories were arranged in five broad groupings: (1) Dispositions, (2) Temporary conditions, (3) Social and reputational aspects, (4) Overt characteristics and appearance, and (5) Terms of limited utility. It is at this higher level of abstraction that the convergences between the present category system and the earlier ones reviewed in Table 3 should be particularly apparent. In addition to the full category labels (e.g. emotions, moods, and cognitions are all considered experiential states), we give an example term and its approximate English translation. These translations point to an interesting and important difference between English and German: Many German words are compounds, consisting of more than one basic concept. In English, such compounds would not be lexicalized as single words and would therefore not appear even in unabridged dictionaries. For example, complex but useful concepts, such as 'understands and appreciates art' (*kunstverständlich*) and 'acts as one would expect a friend to act' (*freundschaftlich*), appeared frequently in our masterpool of adjectives but only infrequently in dictionary-based taxonomies in English.

Classes of person descriptors: a prototype conception

Most efforts to classify person-descriptive terms or statements have employed mutually exclusive categories. For example, in their classification of nearly 18,000 English person descriptors Allport and Odbert assigned each term to one of their four categories. Both Allport and Odbert's and Norman's classifications are examples of a classical or Aristotelian conception of categories (Smith and Medin, 1981). In a classical conception, the entire corpus of person-descriptive terms is divided into discrete classes, such as traits, states, and activities. Each term is assigned to only one class, and within each class all members are logically equivalent. This idealized conception, however, applies neither to natural categories in general (Rosch, 1978) nor to person-descriptive terms in particular (Allen and Potkay, 1981). An inspection of Allport and Odbert's and of Norman's classifications quickly shows that the categories overlap and have fuzzy boundaries. Indeed, Allport and Odbert themselves noted that some of the terms could have been classified into more than one category, especially those assigned to the 'Trait' and the 'State and activity' categories.

This observation has led some researchers (Allen and Potkay, 1981; 1983) to argue that distinctions between classes of person descriptors are arbitrary and should be abolished. There are, however, several alternatives to such an extreme conclusion. One solution comes from recent research on emotion descriptors. For example, Ortony, Clore and Foss (1987) called the boundary between the state and trait classes 'murky' (p. 354) and argued that descriptors 'vary in what might be called their "dispositional potential"'. Some words refer only to trait-like dispositions and resist any state reading at all (e.g. 'competent,' and 'trustworthy') ... Other words can never be given trait readings (e.g. 'gratified'), and yet others are ambiguous, having both a trait reading and a state reading (e.g. 'happy')' (1987, p. 350). Ortony *et al.* therefore divided the domain into terms that (a) clearly refer to states, (b) clearly refer to dispositions (called 'frames of mind'), and (c) borderline cases (called 'state-like conditions').

However, the 'unclear' cases in this intermediate class create a problem only if one insists on classical definitions in terms of necessary and sufficient attributes.

Table 4. Classes of person characteristics in German: The category system with an example for each category, and the number (and percentage in the total pool) of German adjectives assigned to that category by the majority of the judges

(1) Dispositions		601 (12%)	
(1a) Temperament and character traits	Dickköpfig (bullheaded)	286 (6%)	
(1b) Abilities, talents, or their absence	Kunstverständnis (art-understanding)	125 (2%)	
(2) Temporary conditions		562 (11%)	
(2a) Experiential states: Emotions, moods, and cognitions	Stinkwütend (stinking-furious)	276 (5%)	
(2b) Physical and bodily states	Kurzatmig (out-of-breath)	72 (1%)	
(2c) Behavioral states: Observable activities	Hastig (in-a-rush)	49 (1%)	
(3) Social and reputational aspects		885 (17%)	
(3a) Roles and relationships	Freundschaftlich (acts as would a friend)	31 (0.6%)	
(3b) Social effects: Reactions of others	Herzgewinnend (wins others' hearts)	21 (0.4%)	
(3c) Pure evaluations	Hassenswert (merits hatred)	464 (9%)	
(3d) Attitudes and worldviews	Erzkonservativ (archconservative)	116 (2%)	
(4) Overt characteristics and appearance		198 (4%)	
(4a) Anatomy, constitution, and morphology	Vollbusig (busty)	87 (2%)	
(4b) Appearance, looks, and deportment	Modisch (fashionable)	60 (1%)	
(5) Terms of limited utility		151 (3%)	
(5a) Context-specific or technical	Wehrpflichtig (has to serve in military)	117 (2%)	
(5b) Metaphorical, vague, or outmoded	Töchterreich (rich in daughters)	10 (0.2%)	
Number of terms receiving majority classifications (at least 60% of the judges)		2397 (47%)	
Number of terms in the initial pool		5092 (100%)	1715 (34%)

An alternative view is provided by a prototype conception where each category is defined in terms of its clear cases rather than its boundaries (Rosch, 1978). This conception does not require that categories of person descriptors be discrete and that all their instances be clearly defined. Instead, each category is represented by its most prototypical exemplars; category membership does not need to be defined strictly but is a matter of degree.

Chaplin *et al.* (1988) have applied such a prototype conception to three classes of person descriptors, namely traits, states, and activities. First, although Chaplin *et al.* found that the discrete classification of some descriptors (e.g. energetic, suspicious, antagonistic) was difficult, the classification of most instances was quite clear. Second, as predicted by the prototype conception, the core of each category could be differentiated from that of the others by a set of attributes. For example, prototypical *states* were seen as temporary, brief, and externally caused; in contrast, prototypical *traits* were seen as stable, long-lasting, and internally caused, and needed to be observed more frequently and across a wider range of situations than states before they were attributed. And third, Chaplin *et al.* provided evidence for the convergence of several measures of prototypicality. In particular, the rated prototypicality of a term vis-a-vis a category was highly correlated with the percentage of judges classifying the term into that category, both in a discrete classification task (i.e. either state or trait) and in a multiple classification task (i.e. state, trait, both, or neither). Therefore, if a large number of stimuli have to be classified into a large number of alternative categories, the prototypicality of a descriptor vis-a-vis the set of categories can be indexed by the percentage of judges classifying the term in the category. Overall, then, Chaplin *et al.*'s findings demonstrate that the conceptual distinctions made by Allport and Odbert (1936) and by Norman (1967) are rooted in a common understanding of personality and should be conceptualized as prototype categories with fuzzy boundaries.

Method

We examined the largest of the three personality-relevant word classes, the personality adjectives. Moreover, because most taxonomers have focused almost exclusively on adjectives, cross-language comparisons are possible only for that class of descriptors.

Subjects

A total of 10 judges, including advanced graduate students in psychology, a Ph.D. psychologist, and a psychological-technical assistant, were recruited for the classification task.

Materials and design

The rating task was structured so as to permit the investigation of three reliability indices: interjudge agreement, temporal stability, and validity against an expert classification. The total pool of 5,092 adjectives was divided into 51 sets of roughly 100 adjectives each. To assess the temporal stability of the judgments, two items were randomly selected from each of the 51 sets of terms. These 100 retest adjectives were assembled in a separate rating booklet and presented for a second rating after the judge had completed the ratings of 5,092 terms.

The judges received extensive written instructions providing a conceptual definition

of each category in the taxonomy and a discussion of the differences among the categories. At the end of the written instructions, the judges received a list of 140 example adjectives that served to exemplify the prototypical core of each of the categories. These 140 terms had been classified jointly and unanimously by the three authors and were interspersed among the 5,092 adjectives. If this expert classification is taken as a standard for the way the category system was intended to be used, the validity with which the judges used each category can be assessed by correlating the expert classifications with the classifications by the 10 judges.

Initial judgments: synonym, clarity, and personality relevance

The judges were given instructions and a response sheet that organized the task into a series of steps. In order to force the judges to consider seriously the meaning of the adjective they were about to classify, they were asked to first generate a synonym for the adjective or, if they could not think of one, to write down a short definition. The judges then rated their familiarity with the meaning of the adjective on a 3-step scale ranging from '1' (the meaning of the word is not clear enough for me to complete the subsequent ratings), '2' (the meaning of the word became clear to me only after giving it some thought), to '3' (the meaning of the word is fairly clear to me). If the meaning was too unclear to provide any further ratings, the judge checked '1' on the Clarity of meaning scale and moved on to the next adjective. If the meaning was clear enough, the judge next rated the Personality relevance of the adjective, defined by the question 'Can you imagine this term being used for the description of an individual, or for the description of an individual's experience, behaviour, or appearance?' If the judge responded '1' (impossible to imagine) or '2' (unusual; possible to imagine it only under certain conditions), the term was considered not clearly personality relevant. Only after ratings of '3' on personality relevance (easy to imagine personality use) was the judge to move on to the classification task.

Classification task

Subjects were given definitions of each category and were trained in the use of the category system.⁴ For each adjective that passed the Clarity of meaning and Personality relevance tests, the judge considered whether the adjective fit in any of the categories. Category (5) 'Terms of limited utility' was not, strictly speaking, a content category. Rather, it was defined specifically to include terms relevant for personality descriptions but limited in use to specific contexts, population groups, or content domains. Specifically, category (5a) included terms that apply only to specific groups (e.g., 'wehrpflichtig' can be applied only to men in the age range of 19 to 29), describe sexual orientations and peculiarities or, most generally, refer to medical, psychological, or disease terminology. Category (5b) included adjectives whose meaning or implication for personality is not readily apparent, vague, or must be inferred by metaphorical thinking. Also included here were unusual, currently unused, and outmoded terms. We also provided the judges with an exclusion category, to be used when an adjective did not fit any of the content categories. Finally, the prototype

⁴ A description of the category system, including detailed rating instructions and examples, is available from the authors.

model suggests that some adjectives cannot be classified in just one category. We therefore permitted *double classifications*; that is, judges could classify the same adjective into more than one category if two categories appeared equally plausible or relevant. In general, however, we encouraged the judges to classify each adjective into the one ‘best fitting’ category.

Procedure

The judges received the booklets containing subsets of roughly 100 adjectives one at a time and completed each at their own pace and convenience. The task was very time-consuming. Even after training and considerable practice, the judges needed more than two hours to classify a set of 100 adjectives. To ensure that they would work as carefully as possible, the judges were given no time limit. As consequence, the categorization of the entire pool of adjectives by each of the 10 judges required almost four years. Note that the second administration of the 100 retest adjectives occurred at the end of this period. Thus, the average term on the retest list was rated again by the judge after a retest interval of about *two* years. After two years—and about 5,000 adjectives later—memory of the previous ratings must be considered all but nonexistent.

Results

Synonym, clarity of meaning, and personality relevance

The mean of the judges’ mean ratings on the 3-step Clarity of meaning scale was

2.8 (SD = 0.4); 57% of the adjectives received a mean rating of 3.0, and in only 8.8% of all judgments was the adjective eliminated because the judge considered its meaning as too unclear. The coefficient Alpha reliability of the mean ratings was 0.90, indicating substantial agreement among the judges. The two-year retest stability of the mean ratings, computed as a correlation across the 100 adjectives administered twice, was even higher, 0.96.

To examine the ‘behavioral’ validity of the Clarity of meaning rating, we analyzed the synonyms and definitions listed by the judges. Overall, the number of judges able to list a synonym for the average term was 7.3, and another 1.7 judges were able to at least generate a definition of the term. In other words, for the average term 9 judges listed either a synonym or a definition, implying that the judges felt quite familiar with most of the words. The correlation between the overall behavioral index of familiarity (i.e. number of judges who could list a synonym or definition) and the mean rating on the Clarity scale was 0.96, providing considerable evidence for convergent validity.

The mean rating on the Personality relevance scale was 2.6 (SD = 0.5); 40% of the adjectives received a mean rating of 3.0. The coefficient Alpha reliability of the mean ratings was 0.87, again indicating substantial interjudge agreement, and the two-year retest stability of the mean ratings (across 100 adjectives) was an impressive 0.92. In all, the initial ratings of Clarity of meaning and Personality relevance were reliable across both judges and time, and—for the Clarity of meaning scale—correlated highly with a behavioral criterion of word familiarity.

Reliability of the category system

To obtain a measure of the degree to which an adjective fit in a particular taxonomic category, we computed a prototypicality score reflecting the number of judges who classified the adjective in that category (cf. Chaplin *et al.*, 1988). The quality of the category system, and the reliability with which the judges used each of the categories, was evaluated in terms of (a) the internal consistency (Coefficient Alpha) of these prototypicality scores across all 5,092 terms, (b) the stability of the prototype scores assessed at two different times for a subsample of 100 terms, and (c) their correlation with the expert classifications across the 140 terms for which both kinds of classifications were available. All three types of indices are presented in Table 5, separately for the five superordinate, and the 13 subordinate categories.

Table 5. Categorizing adjectives into classes of person characteristics: Reliability of the prototypicality scores

Abbreviated category label	Interjudge agreement	Stability over 2 years	Correlation with experts
(1) Dispositions	0.84	0.95	0.85
(1a) Temperament and character	0.78	0.87	0.82
(1b) Abilities and talents	0.82	0.93	0.90
(2) Temporary conditions	0.86	0.90	0.87
(2a) Experiential states	0.87	0.89	0.83
(2b) Physical and bodily states	0.88	0.90	0.96
(2c) Observable activities	0.63	0.51	0.84
(3) Social and reputational aspects	0.83	0.88	0.84
(3a) Roles and relationships	0.70	0.48	0.94
(3b) Social effects	0.63	0.60	0.87
(3c) Pure evaluations	0.82	0.88	0.82
(3d) Attitudes and worldviews	0.90	0.95	0.89
(4) Overt characteristics and appearance	0.91	0.91	0.92
(4a) Anatomy and constitution	0.90	0.96	0.95
(4b) Appearance, looks, deportment	0.81	0.79	0.88
(5) Terms of limited utility	0.75	0.82	0.70
(5a) Context-specific or technical	0.80	0.84	0.80
(5b) Metaphorical, vague, outmoded	0.53	0.63	0.72

Note: An adjective's prototype score for a category is the number of judges who classified the adjective in that category. Interjudge agreement is indexed by the Coefficient Alpha based on the intercorrelations among the 10 judges across all 5,092 terms. Stability is the correlation of the prototype scores for 100 terms administered twice. The correlation with the expert criterion classification was computed across 140 terms classified unanimously by the three authors prior to data collection.

The Coefficient Alpha reliabilities varied from 0.75 to 0.91 for the superordinate categories, and from 0.53 to 0.90 for the subordinate categories. In general, these coefficients imply a substantial degree of generalizability of these classification scores to another sample of judges. The Alphas were slightly higher for the superordinate than for the subordinate categories. This finding suggests that the finer and more differentiated distinctions required at the level of the subordinate categories made the judgment task more difficult and evoked classifications that were more idiosyncratic. We will examine this possibility in more detail later. The most observable category in the taxonomy—Overt characteristics and appearance (4)—elicited the most consen-

sus. The lowest Alpha reliabilities, in contrast, were observed for categories that were not used frequently by the judges. In particular, (5b) Metaphorical, vague, and outmoded terms (Alpha = 0.53), (3b) Social effects: Reactions of others (0.63), (3a) Roles and relationships (0.70), (2c) Behavioral states: Observable behaviors (0.63), and (4b) Appearance, looks, and deportment (0.81) are the categories that were (in ascending order) used least frequently.

The two-year temporal stability correlations, computed across 100 adjectives administered twice, were higher than we expected. For each of the superordinate categories, the stability coefficients exceeded 0.80, ranging from 0.82 to 0.95. For most of the subordinate categories, the stabilities were generally of similar size; however, the four categories that elicited the least interjudge agreement showed markedly lower, though still significant, levels of temporal stability (r_s of 0.48, 0.51, 0.60, and 0.63).

The last two columns of Table 5 list the point-biserial correlations between the prototypicality scores and the dichotomous classifications made jointly and consensually by the three experts for a subset of 140 terms. These correlations ranged from 0.70 to 0.92 for the superordinate categories, and from 0.72 to 0.96 for the subordinate categories. For each of the substantive categories (i.e. excluding category 5), the correlations with the experts clearly exceeded 0.80. If our 'expert' classification can be taken as a standard for the way we intended the judges to use the category system, the size of these correlations would suggest that this group of judges came, in most instances, quite close to the ideal.⁵ In conclusion, the prototypicality scores were found to generalize both across our sample of judges and over a two-year time interval, and they converged with the intended use defined by the classifications of an expert panel.

Prototypes for the categories

Among the total of 53,436 category assignments made by the 10 judges, 4.7% were multiple classifications. By far the largest number of double classifications occurred for terms at the boundary between traits and states, represented here by Character and temperament traits (1a) and Experiential states (2a). Some of the terms receiving double classifications in these two categories are *zuversichtlich* (optimistic), *springlebendig* (jumpingly lively), *eifersüchtig* (jealous), *leidenschaftlich* (passionate), *gelassen* (relaxed), and *zartfühlend* (feeling tenderly). There were even a few terms that received a sufficient number of double classifications so as to qualify as prototypical members of both categories (e.g. *ängstlich* [anxious]).

The 'unclassifiable' category was used very infrequently; on the average, the judges were able to assign the adjective to a content category in all but 0.7% of the cases. These findings suggest that for the vast majority of the adjectives each judge was able to select an appropriate or 'best fitting' category from those included in the taxonomy. In other words, the taxonomy was sufficiently exhaustive to accommodate almost all the adjectives. The existence of multiple classifications shows that the

⁵ We should note, however, that the judges received these 140 adjectives, and their expert classifications, as a way to exemplify the categories. Thus, it is possible that these high correlations reflect, in part, the judges' conscientious and thorough adherence to the appendix of the instruction manual. However, after the initial training and some hundred judgments, the judges seldom needed to refer back to the manual. Given that the 140 terms had been interspersed among the 5,092 adjectives each judge had to rate, a simple memory effect seems not particularly likely although we cannot rule it out.

prototype conception was needed but the moderate number of double classifications suggests that the categories are nonetheless distinguishable.

In our conception of the category system, we defined category membership in a graded, continuous way. However, for various purposes, such as comparisons with earlier, 'all or none' classification systems, an analysis of the prototypical core of each category can be useful. A truly prototypical example of a category should be classified in that category by at least the majority of the judges, that is, by at least six of our 10 judges. Using that 'at least 60%' cut-off rule, 2,397 of the 5,092 adjectives (47%) would be considered highly prototypical members of one of the five superordinate categories.⁶ The finding that almost *half* of all adjectives could be classified uniquely at the superordinate level suggests that the category definitions were intuitively sensible and capture meaningful distinctions inherent in the German lexicon of personality descriptors. Apparently, the classifications reflect stable and consensually shared beliefs about the ways in which personality can be described. Finally, the relatively clear separation of the category prototypes suggests that the taxonomy identifies distinctions in the structure of the German language.

The second to last column in Table 4 shows the breakdown of these 2,397 prototypical terms into the five superordinate categories. The largest category, accounting for 17% of the total pool, was Social and reputational aspects (3), followed by Dispositions (1) with 12% and Temporary conditions (2) with 11%. The remaining two categories were less extensively represented in the German personality lexicon; Overt characteristics and appearance (4) accounted for 4%, and Terms of limited utility (5) for 3% of the total pool.

As we had noted in the reliability analyses, interjudge agreement was lower for the subordinate categories, suggesting that some of the specific distinctions within the superordinate categories were difficult to make. The lower interjudge agreement for the subordinate categories leads to an equal, or almost equal, number of classifications of the same adjective in two different subordinates within the same superordinate category. For example, about half the judges classified *angesehen* (respected), *lästig* (bothersome), and *hochinteressant* (highly interesting) in Social effects (3b), whereas the other half classified these adjectives as Pure evaluations (3c). Obviously, there was nearly unanimous agreement among the judges that these terms refer to Social and reputational aspects (3), but the further distinction within that superordinate category was essentially arbitrary. These terms describe both a social effect *and* an evaluation, and thus belong equally to both categories. However, although these terms are highly prototypical exemplars of the superordinate category, they do not fit particularly well in just one of the subordinate categories. There are many examples of this type within the other superordinate categories as well. For example, the overlap between Temperament and character traits (1a) and Abilities and talents (1b) is illustrated by terms such as *vielseitig* (versatile), *listig* (cunning), *aufgeweckt* (alert), *einfühlsam* (empathic), and *unbeholfen* (clumsy). In general, these terms refer to traits of temperament (e.g. alert) or character (e.g. cunning) that require a particular capacity, ability, skill, or talent.

The implications of this effect for the distribution of the prototypicality scores is shown in the last column of Table 4. According to the 'at least 60%' cut-off

⁶ If we only slightly relaxed this 'majority' criterion, requiring classification in that category by only half the judges (i.e. 5 out of 10), the total number of terms considered prototypical of the superordinate categories would rise to 3,257, almost two-thirds of the total pool of adjectives studied here.

rule, only 1,714 of the 5,092 terms would be considered highly prototypical exemplars of the 13 subordinate categories—that is, substantially fewer unique classifications (34%) than at the superordinate level (47%). This finding can be described by the bandwidth—fidelity trade-off in personality categories (see Hampson *et al.*, 1986; John *et al.*, 1990): The advantage of narrow categories is that they provide more detailed informative classifications than broader categories; their disadvantage is that they overlap more with each other and have fewer uniquely prototypical instances than do broad categories. In the present study, much distinctiveness is sacrificed in exchange for the increase in informativeness at the subordinate level, and for many purposes one would want to use only the five superordinate categories.

Nonetheless, the distribution of the 1,714 ‘prototypical’ adjectives across the 13 subordinate categories was similar to that of the superordinate categories. Category (3c) Pure evaluations was the largest subordinate (9%), followed by (1a) Temperament and character traits (6%) and (2a) Experiential states (5%). In contrast, (3a) Roles and relationships and (3b) Social effects, even when combined, barely reached the 1% mark. The two categories for Overt characteristics (4a and 4b), and those for Physical states and Behavioral states (2b and 2c), were also fairly small; smaller actually than the two categories studied here for the first time, (1b) Abilities, talents, or their absence and (3d) Attitudes and worldviews.

In interpreting these percentages, we should keep in mind the arbitrariness of the prototype assignments on which they are based. In Table 6 we report for each category the number of adjectives at each level of prototypicality, ranging from ‘0’ (none of the judges classified that term in that category) to ‘10’ (all the judges classified the term in that category). In Table 6, values of 6 and above (i.e. majority classifications) have been set in italics. Note, however, that there is nothing magical about the number 6. The distribution of the values is entirely continuous, and there is no obvious break in the distribution that would single out any one point as the natural cut-off.

Category frequencies for German and American terms

Comparisons with the category frequencies in Norman’s (1967) classification are difficult because of several differences between the two category systems. For one, Norman grouped his terms within content categories into ‘prime,’ ‘difficult’ (i.e. unfamiliar), and ‘slangy and quaint’ terms. Moreover, a full 47% of Norman’s terms were assigned to exclusion categories because they were too obscure, ambiguous, or vague. In contrast, our analyses of the clarity of the word meaning and of the number of synonyms generated for the word showed that word familiarity (or difficulty) had, at best, a minor effect on our classifications. In contrast to English with its massive and largely unused vocabulary, German contains relatively few words whose meaning would not be understood by a well-educated speaker. Indeed, although our Master List had not been culled for difficult and ‘outdated’ words, less than 10% of the terms were rejected because their meaning was not clear enough.

Given the differential percentages of terms eliminated as a function of word familiarity alone, the only sensible comparisons are those relative to the other categories of descriptors. First, Norman found, as did we, a similar number of stable traits (or dispositions) and of states and activities (or conditions), each amounting to about 16% of Norman’s total pool of 18,125 terms. Moreover, the numbers for the Anatomy and Appearance terms (which Norman treated as one category) were low and about

Table 6. Distributions of the prototypicality scores for all 13 categories: The number of adjectives receiving each prototypicality score

Abbreviated category label	Prototypicality score										
	0	1	2	3	4	5	6	7	8	9	10
1a Traits	2503	1039	517	334	254	159	140	76	50	20	0
1b Talents	3728	743	260	118	83	39	34	26	29	26	10
2a Experiential states	3462	678	327	164	95	90	74	55	51	55	41
2b Physical states	4663	201	59	37	41	19	20	17	17	15	3
2c Behavioral states	3276	981	424	210	105	47	29	13	6	1	0
3a Roles	4184	599	168	61	33	16	10	9	7	3	2
3b Effects	3983	707	224	78	53	26	11	6	2	1	1
3c Evaluations	2333	976	558	349	257	155	150	119	105	64	26
3d Attitudes	4447	298	104	45	36	46	21	21	27	24	23
4a Anatomy	4658	182	72	39	30	24	24	12	22	24	5
4b Appearance	4318	411	143	76	54	30	20	16	9	10	5
5a Technical terms	3678	834	241	112	60	50	33	29	28	17	10
5b Vague terms	3980	719	251	90	30	12	7	2	1	0	0

Note: The prototypicality score of an adjective for each category is the sum of the 10 judges assigning the adjective to that category, including double classifications (i.e. 4.7% of all classifications). Scores of 6 and above mean that a majority of at least 60% of the judges assigned the adjective to that category.

the same in the two studies. However, whereas we found more Pure evaluations than either Dispositions or Temporary conditions, Norman reported a far smaller number of Evaluative terms (4%). Conversely, Norman found a much larger number of Social roles and of Social effects (8%) than did we.

These differences are probably due to the different position the Evaluative terms were assigned in the two taxonomies. Norman considered the Evaluative terms as an exclusion category, to be used sparingly. In contrast, we grouped the Pure evaluations along the other social and reputational aspects of individuals. As shown by an analysis of terms with high classifications in more than one category, our Pure evaluation category (3c) overlapped considerably with its neighboring category Social effects (3b), as well as with Temperament and character traits (1a), Appearance (4b), and Abilities and talents (1b). Obviously, many of the descriptors in these categories are highly evaluative, and the German judges classified as 'purely evaluative' many of the characteristics they considered extreme and unusual (e.g. heroic, brilliant, stupid, mean), as well as many of the terms related to moral values (e.g. amoral, improper, guilty), to interpersonal pleasantness (e.g. unpleasant, loving, desirable), and to appearance (e.g. picture-pretty, ugly as death). Norman's team, in contrast, classified some of these terms into substantive categories, such as traits, social roles or effects, and appearance, leading to a more extensive representation of these categories and, at the same time, keeping the Evaluative category to a more limited size.

Whereas these differences probably resulted from different thresholds for classifying descriptors as 'purely evaluative,' we also found substantially more Experiential

states than Behavioral states (5% vs. 1%) among the German adjectives, where Norman found the opposite pattern (2% vs. 3% among the 'prime' terms). One explanation for this difference might lie in the lexicalization rules in English and German. In English, the category of Activity-descriptive adjectives consists, almost exclusively, of gerunds, such as carousing, ranting, and discussing (see Table 1 in Chaplin *et al.*, 1988). The same type of verb form, ending in *-nd* in German, was found among the most prototypical activity descriptors in our list, for example, *nörgelnd* (complaining), *planend* (planning), and *zögernd* (hesitating). However, such gerunds are listed only infrequently in the German dictionary because they can be formed from every verb, and listing them as a separate entry would just create an additional, redundant listing.

Whereas in English almost all activity descriptors are gerunds ending in *-ing*, only 16 of our 49 adjectives classified as prototypical activity descriptors ended in the suffix *-nd*. An additional 13 activity adjectives were formed not from verbs but from nouns with the suffix *-ig*, such as *leichtfüßig* (light-footed) and *mutwillig* (intentional), and from various other combinations of nouns or verbs with particular suffixes, such as *selbstmörderisch* (suicidal) and *verächtlich* (derogatory). In conclusion, the class of lexicalized activity descriptors is probably smaller in German than in English because it draws upon the large class of verbs only to a limited extent; instead, a relatively larger proportion of the German activity-descriptive adjectives are formed from nouns.

These initial comparisons with Norman's English-language taxonomy have only scratched the surface of the kind and scope of cross-language analyses that eventually can be performed. Many of our interpretative difficulties arose from differences in procedures and from insufficiently explicated strategies and biases on the part of the judges in different studies. Nonetheless, we have been able to highlight a few interesting differences between English and German, particularly in the lexicalization rules of these two closely related languages. In future taxonomic research, however, we need to establish greater uniformity and objectivity in our procedures if findings and knowledge are to be cumulative and eventually comparable across languages.

CONCLUSIONS AND IMPLICATIONS

The major goal of this article was to replicate and extend taxonomic analyses and procedures that have so far been applied only to American English and to Dutch, and to thus lay the groundwork for systematic and detailed taxonomies of German personality terms. The present studies have resulted in unabridged Master Lists of three types of personality-relevant German words: adjectives, type nouns, and attribute nouns. The latter two types now need to be given the same intense research attention that so far has been limited to the adjective class (e.g. De Raad and Hoskens, 1990). On the basis of our analysis of the German type and attribute nouns, we have formulated a number of hypotheses about the function and uses of these descriptors, hypotheses that now should be studied more systematically.

In the second part of the article, we clarified and elaborated a number of basic, conceptual distinctions among ways of describing personality in terms of an extensive and carefully defined category system. We have now tested this category system with a large number of personality adjectives and with 10 judges, over a rather

long period of time. Our analyses suggest that the category system captured most of the content of the personality lexicon in German, and represented these contents in a way that the judges found intuitively meaningful and allowed them to make judgments that were consensually shared and stable over a considerable amount of time.

Finally, this work has been motivated by the need to establish common and standard procedures for future taxonomic research on other, non-germanic languages. In developing our procedures, we have built on, and tried to integrate, the knowledge that has accumulated over the past 50 years since Allport and Odbert (1936) and Cattell (1943) started it all. We hope, therefore, that the present paper can serve as an outline, or blue print, for how one might go about starting a taxonomy 'from scratch' in another language, and that it might help stimulate researchers, in Southern and Eastern European countries in particular, to begin a taxonomy project in their own language.

Faced with Cattell's (1943) legacy, Norman (1967) was the first to recognize the need to ensure the replicability of the basic taxonomic classification we have reported here in our second study. Norman therefore assembled a team of four judges but, unfortunately, reported later only the dichotomous, either-or classifications of all of his terms. The Dutch taxonomy team, however, developed numerous new procedures to examine and ensure the replicability of every stage in the project.

In classification tasks of the kind used in the present research, judges show tremendous differences in the number of items they are willing to assign to any particular category. For example, in our first study, the two judges who examined the same part of the dictionary often differed substantially in the number of terms they considered personality relevant. In our second study, we found, as did Allport and Odbert 54 years earlier, that different judges have different and often quite pronounced preferences for particular categories. For example, there was an enormous range in the numbers of descriptors individual judges would assign to the Trait and Activity categories. The most extreme judge in the current sample saw personality as a product of social and behavioral learning and thus classified only 35 adjectives in the trait category, as compared with the 1600 trait classifications made by the most extremely trait-orientated judge in the present sample.

Thus, our finding that the group means were generalizable across judges and stable over time should not be taken to mean that there were no consistent individual differences among the raters. After all, coefficient Alpha estimates pertain to the generalizability of group means, not to the agreement between individual judges. For example, an Alpha of 0.80 for a group of 10 judges is equivalent to a correlation of about 0.30 between pairs of individual judges. In other words, if there is only one judge (or investigator) to make the judgments, classifications, or necessary taxonomic decisions, those judgments will probably correlate not much higher than 0.30 with those made by another judge (or investigator) trying to replicate the earlier effort. In other words, aggregation was central, and necessary, at every step of our project, to reach the levels of reliability, retest reliability, and expert agreement obtained in the present research. The many decisions and subjective judgments made during a taxonomic project must be based, we believe, on a panel of independent judges whose performance is evaluated and monitored.

This point is worth some emphasis because, with the exception of the Dutch project, all previous taxonomers have made, and continue to make, their most crucial

decisions and judgments by themselves. Moreover, if these decisions are made by a 'team,' that team does not seem to function as a group of independent judges whose agreement and reliability can be tallied and evaluated. Indeed, the 'one person, one taxonomy' habit seems to persist even into analyses of the Big Five, where each investigator (or pair of investigators) has to pick out their own, idiosyncratically selected, set of variables.

In our earlier review of the taxonomy literature (John *et al.*, 1988), we noted, with some surprise, the remarkable differences even among several American taxonomies that all began with essentially the same set of 1,710 descriptors (Goldberg, 1982). Indeed, convergence among different taxonomic systems has been achieved, if at all, only at the level of the broadest dimensions (the Big Five), and convergence on a set of more specific 'middle-level' categories or facets (Briggs, 1989) is nowhere in sight. Nevertheless, replicability is the foremost, and most basic, standard of science, and although some taxonomers hold that taxonomy is as much an art as it is a science, we believe we need more science and less art in the field at the present time. It is about time that we learn some lessons from the history of Cattell's (e.g. 1943) daring taxonomic ventures during the pre-computer dark ages of taxonomy construction.

REFERENCES

- Allen, B. P. and Potkay, C. R. (1981). 'On the arbitrary distinction between states and traits', *Journal of Personality and Social Psychology*, **41**: 916-928.
- Allen, B. P. and Potkay, C. R. (1983). 'Just as arbitrary as ever: Comments on Zuckerman's rejoinder', *Journal of Personality and Social Psychology*, **44**: 1087-1089.
- Allport, G. W. and Odbert, H. S. (1936). 'Trait-names: A psycho-lexical study', *Psychological Monographs*, **47**, No. 211.
- Angleitner, A., John, O. P. and Löhr, F.-J. (1986). 'It's *what* you ask and *how* you ask it: An itemmetric analysis of personality questionnaires'. In: Angleitner, A. and Wiggins, J. S. (Eds), *Personality Assessment via Questionnaires: Current Issues in Theory and Measurement*, Springer-Verlag, Berlin, pp. 61-108.
- Anglin, J. M. (1977). *Word, Object, and Conceptual Development*, Norton, New York, NY.
- Baumgarten, F. (1933). 'Die Charaktereigenschaften' [The character traits]. In: *Beiträge zur Charakter-und Persönlichkeitsforschung* [Contributions to Research in Character and Personality] (Whole No. 1), A. Francke, Bern, Switzerland.
- Bond, M. H. (1979). 'Dimensions used in perceiving peers: Cross-cultural comparisons of Hong Kong, Japanese, American and Filipino university students', *International Journal of Psychology*, **14**: 47-56.
- Bond, M. H. (1983). 'Linking person perception dimensions to behavioral intention dimensions: The Chinese connection', *Journal of Cross-Cultural Psychology*, **14**: 41-63.
- Bond, M. H. and Forgas, J. P. (1984). 'Linking person perception to behavior intention across cultures: The role of cultural collectivism', *Journal of Cross-Cultural Psychology*, **15**: 337-352.
- Bond, M. H., Nakazato, H. and Shiraishi, D. (1975). 'Universality and distinctiveness in dimensions of Japanese person perception', *Journal of Cross-Cultural Psychology*, **6**: 346-357.
- Borkenau, P. (1990). 'Traits as ideal-based and goal-derived social categories', *Journal of Personality and Social Psychology*, **58**: 381-396.
- Briggs, S. R. (1989). 'The optimal level of measurement for personality constructs'. In: Buss, D. M. and Cantor, N. (Eds), *Personality Psychology: Recent Trends and Emerging Directions*, Springer-Verlag, New York, NY, pp. 246-260.
- Brokken, F. B. (1978). *The Language of Personality*, Krips, Meppel, The Netherlands.
- Bromley, D. B. (1977). *Personality Description in Ordinary Language*, Wiley, London, UK.

- Buss, D. M. and Craik, K. H. (1983). 'The act frequency approach to personality', *Psychological Review*, **90**: 105–126.
- Cantor, N. and Mischel, W. (1979). 'Prototypes in person perception'. In: Berkowitz, L. (Ed), *Advances in Experimental Social Psychology*, Vol. 12, Academic Press, New York, NY, pp. 2–52.
- Cattell, R. B. (1943). 'The description of personality: Basic traits resolved into clusters', *Journal of Abnormal and Social Psychology*, **38**: 476–506.
- Chaplin, W. F., John, O. P. and Goldberg, L. R. (1988). 'Conceptions of states and traits: Dimensional attributes with ideals as prototypes', *Journal of Personality and Social Psychology*, **54**: 541–557.
- Church, T. A. and Katigbak, M. S. (1989). 'Internal, external, and self-report structure of personality in a non-western culture: An investigation of cross-language and cross-cultural generalizability', *Journal of Personality and Social Psychology*, **57**: 857–872.
- De Raad, B. (1984). 'Person-talk in everyday situations'. In: Bonarius, H., Van Heck, G. and Smid, N. (Eds), *Personality Psychology in Europe: Theoretical and Empirical Developments*, Swets & Zeitlinger, Lisse, The Netherlands, pp. 31–44.
- De Raad, B. and Hoskens, M. (1990). 'Personality-descriptive nouns', *European Journal of Personality*, **4**: this issue.
- De Raad, B., Mulder, E., Kloosterman, K. and Hofstee, W. K. B. (1988). 'Personality-descriptive verbs', *European Journal of Personality*, **2**: 81–96.
- Dixon, R. M. (1977). 'Where have all the adjectives gone?', *Studies in Language*, **1**: 19–80.
- Fiske, S. T. and Cox, M. G. (1979). 'Person concepts: The effects of target familiarity and descriptive purpose on the process of describing others', *Journal of Personality*, **47**: 136–161.
- Goldberg, L. R. (1981). 'Language and individual differences: The search for universals in personality lexicons'. In: Wheeler, L. (Ed), *Review of Personality and Social Psychology*, Vol. 2, Sage, Beverly Hills, CA, pp. 141–165.
- Goldberg, L. R. (1982). 'From Ace to Zombie: Some explorations in the language of personality'. In: Spielberger, C. D. and Butcher, N. (Eds), *Advances in Personality Assessment*, Vol. 1, Erlbaum, Hillsdale, NJ, pp. 203–234.
- Gough, H. G. and Heilbrun, A. B., Jr. (1965). *Adjective Check List Manual* (rev. ed. 1985), Consulting Psychologists Press, Palo Alto, CA.
- Gross, D., Fischer, U. and Miller, G. A. (1989). 'The organization of adjectival meanings', *Journal of Memory and Language*, **28**: 92–106.
- Guilford, J. P. (1959). *Personality*, McGraw-Hill, New York, NY.
- Guthrie, G. M. and Bennett, A. B., Jr. (1971). 'Cultural differences in implicit personality theory', *International Journal of Psychology*, **6**: 305–312.
- Hampson, S. E., John, O. P. and Goldberg, L. R. (1986). 'Category breadth and hierarchical structure in personality: Studies of asymmetries in judgments of trait implications', *Journal of Personality and Social Psychology*, **51**: 37–54.
- Hogan, R. (1983). 'A socioanalytic theory of personality'. In: Page, M. (Ed), *Nebraska Symposium on Motivation, 1982: Personality—Current Theory and Research*, University of Nebraska Press, Lincoln, NE, pp. 55–89.
- Janke, W. and Debus, G. (1978). *Die Eigenschaftswörterliste (EWL). Eine mehrdimensionale Methode zur Beschreibung von Aspekten des Befindens* [The Trait Term List (EWL). A Multi-dimensional Method for the Description of Aspects of Feelings], Hogrefe, Göttingen, FRG.
- John, O. P. (1982). *German Adjective List (GAL)*, Unpublished assessment instrument, University of Bielefeld, Abteilung für Experimentelle und Angewandte Psychologie, FRG.
- John, O. P. (1989). 'Towards a taxonomy of personality descriptors'. In: Buss, D. M. and Cantor, N. (Eds), *Personality Psychology: Recent Trends and Emerging Directions*, Springer-Verlag, New York, NY, pp. 261–271.
- John, O. P. (1990). 'The "Big Five" factor taxonomy: Dimensions of personality in the natural language and in questionnaires'. In: Pervin, L. (Ed), *Handbook of Personality Theory and Research*, Guilford, New York, NY, pp. 66–100.
- John, O. P., Angleitner, A. and Ostendorf, F. (1988). 'The lexical approach to personality: A historical review of trait taxonomic research', *European Journal of Personality*, **2**: 171–203.
- John, O. P., Goldberg, L. R. and Angleitner, A. (1984). 'Better than the alphabet: Taxonomies of Personality-descriptive terms in English, Dutch, and German. In: Bonarius, H., Van

- Heck, G. and Smid, N. (Eds), *Personality Psychology in Europe: Theoretical and Empirical Developments*, Swets & Zeitlinger, Lisse, The Netherlands, pp. 83–100.
- John, O. P., Hampson, S. E. and Goldberg, L. R. (1990). 'Is there a basic level of personality description?', Manuscript submitted for publication.
- Livesley, W. J. and Bromley, D. B. (1973). *Person Perception in Childhood and Adolescence*, Wiley, London, UK.
- McCrae, R. R. (1989). 'Why I advocate the Five-Factor model: Joint factor analyses of the NEO-PI with other instruments'. In: Buss, D. M. and Cantor, N. (Eds), *Personality Psychology: Recent Trends and Emerging Directions*, Springer-Verlag, New York, NY, pp. 237–245.
- McCrae, R. R. (1990). 'Traits and trait names: How well is Openness represented in natural languages?', *European Journal of Personality*, 4: *this issue*.
- Mummendey, H. E., Mielke, R., Maus, G. and Hesener, B. (1977). *Untersuchungen mit einem mehrdimensionalen Selbsteinschätzungsverfahren* [Studies Using a Multidimensional Self-report Technique], Bielefelder Arbeiten zur Sozialpsychologie (No. 4), University of Bielefeld, FRG.
- Nakazato, H., Bond, M. H. and Shiraishi, D. (1976). 'Dimensions of personality perception: An examination of Norman's hypothesis', *Japanese Journal of Psychology*, 47: 139–148.
- Norman, W. T. (1967). *2,800 Personality Trait Descriptors: Normative Operating Characteristics for a University Population*, Department of Psychology, University of Michigan, MI.
- Ortony, A., Clore, G. L. and Foss, M. A. (1987). 'The referential structure of the affective lexicon', *Cognitive Science*, 11: 341–364.
- Rosch, E. (1978). 'Principles of categorization'. In: Rosch, E. and Lloyd, B. B. (Eds), *Cognition and Categorization*, Erlbaum, Hillsdale, NJ, pp. 27–48.
- Smith, E. E. and Medin, D. L. (1981). *Categories and Concepts*, Harvard University Press, Cambridge, MA.
- Wahrig, G. (1981). *Deutsches Wörterbuch* [German Dictionary], Mosaik Verlag, München, FRG.
- White, G. M. (1980). 'Conceptual universals in interpersonal language', *American Anthropologist*, 82: 759–781.
- Wiggins, J. S. (1979). 'A psychological taxonomy of trait-descriptive terms: The interpersonal domain', *Journal of Personality and Social Psychology*, 37: 395–412.
- Wippich, W. and Bredenkamp, J. (1977). 'Bestimmung der Bildhaftigkeit (I), Konkrettheit (C) und der Bedeutungshaltigkeit (m') von 498 Verben und 400 Adjektiven' [Determinants of imaginativeness, concreteness, and meaning of 498 verbs and 400 adjectives], *Zeitschrift für experimentelle und angewandte Psychologie*, 24: 671–680.
- Yang, K. and Bond, M. H. (1989). 'Exploring implicit personality theories with indigenous or imported constructs: The Chinese case', Unpublished manuscript, Department of Psychology, Chinese University of Hong Kong, Shatin, N.T., Hong Kong.

AUTHOR NOTE

Correspondence concerning this article may be addressed to Alois Angleitner, Abteilung Psychologie, Universität Bielefeld, Universitätsstr. 1, D-4800 Bielefeld 1, Federal Republic of Germany, or to Oliver P. John, Department of Psychology, University of California, Berkeley, CA 94720, USA.

This project was supported in part by a research grant from the University of Bielefeld (OZ-2780) and by a Summer Exchange Fellowship from the German Research Foundation (DFG, AN106/5-1) to Alois Angleitner. Support for Oliver P. John was provided by Grant MH-39077 from the National Institute of Mental Health (U.S.A.) and by Biomedical Research Grants 87-20 and 88-24 (University of California). We would like to thank Jack Block, Peter Borkenau, Jochen Fahrenberg, Lewis R. Goldberg, Willem K. B. Hofstee, Rainer Riemann, and Robert A. Wicklund for their comments on earlier drafts of this paper. This research could have never been completed without the dedication of the individuals who provided the judgments analyzed in this paper. We thank them for the time, care, and patience they devoted to the complicated and arduous tasks we asked them to complete.

RÉSUMÉ

Nous présentons deux études qui portent sur une vaste taxinomie de termes allemands descriptifs de la personnalité. Dans la première étude, tous les adjectifs descriptifs de la personnalité (*cynique* par exemple), les substantifs qui décrivent des types de personnalité (*un cynique* par exemple) et les substantifs signifiant des attributs (*cynisme* par exemple) ont été extraits d'un dictionnaire allemand. Nous avons trouvé que presque la moitié de tous les adjectifs allemands était potentiellement pertinente pour décrire la personnalité. Il ne s'agissait que de 8 pourcent pour les substantifs. De plus, il y avait plus de substantifs pour les attributs que pour les types. Les substantifs pour les types sont apparus plus serpentins, métaphoriques, concrets et imagés. Dans la seconde étude, nous discutons les distinctions conceptuelles essentielles entre les unités des descriptions de la personnalité, nous développons un système de catégorisation basé sur une conception-prototype et nous présentons une classification de 5,092 adjectifs en 3 catégories. Les classifications étaient généralisables sur les deux juges et sur un intervalle de temps de deux ans. Les classifications recouvraient celles faites, à priori, par des experts. Une analyse des noyaux prototypiques des catégories différenciées suggère qu'Évaluations, Tempérament et Traits de caractère ainsi qu'États d'expérience sont le plus largement représentés en Allemand, tandis qu'Effets sociaux, Rôles et relations, et Apparence physique apparaissent le moins fréquemment. Ces résultats, bien qu'en général comparables, diffèrent des résultats obtenus par la taxinomie de Norman (1967) concernant le nombre des termes Évaluatifs et le nombre des termes pour Activité. Nos études nous fournissent des listes détaillées et représentatives de mots allemands pour les traits de personnalité, les humeurs et les émotions, les rôles sociaux, les effets sociaux, les évaluations ainsi que le physique. Ces listes peuvent servir de point de départ à des taxinomies, des analyses dimensionnelles et des instruments de mesure. Nous mettons l'accent sur le besoin de procédures standard pour la réalisation d'études taxinomiques et nous faisons des suggestions pour de futures études d'autres langues.

ZUSAMMENFASSUNG

Zwei Studien zur Entwicklung einer übergreifenden Taxonomie von deutschen Persönlichkeitsbegriffen werden vorgestellt. Zuerst wurden alle persönlichkeitsbeschreibenden Adjektive (zynisch), Typensubstantive (Zyniker) und Dimensionssubstantive (Zynismus) aus einem deutschen Wörterbuch herausgeschrieben. Fast die Hälfte aller deutschen Adjektive war potentiell persönlichkeitsrelevant, verglichen mit nur 8% der Substantive. Die Typensubstantive, die seltener vorkamen als die Dimensionssubstantive, waren eher umgangssprachlich, metaphorisch, konkret und bildhaft, wie die Beispiele 'Grossmaul' und 'Holzkopf' zeigen. In der zweiten Studie wurden verschiedene Einheiten der Persönlichkeitsbeschreibung unterschieden, mit Hilfe eines Kategoriensystemes operationalisiert und auf 5,092 Adjektive angewendet. Die Kategorienzuzuweisungen waren sowohl über die Beurteiler als auch über ein zweijähriges Retestintervall hinweg generalisierbar und stimmten mit einer unabhängig vorgenommenen Expertenbeurteilung überein. Soziale Bewertungen, Temperament- und Charaktereigenschaften und Erfahrungszustände waren die am häufigsten vorkommenden Begriffe; soziale Effekte, Rollen und Beziehungen, und äusserliche Erscheinung waren dagegen eher selten. Diese Befunde sind den von Norman (1967) berichteten im allgemeinen ähnlich, weichen aber in der Häufigkeit von bewertenden Begriffen und Aktivitätsbeschreibungen von ihnen ab. Zur Beschreibung von Persönlichkeitseigenschaften, Stimmungen und Gefühlen, sozialen Rollen und Bewertungen, der Wirkung einer Person auf andere, und der äusserlichen Erscheinung stehen jetzt repräsentative Wortlisten zur Verfügung, die die Grundlage für Taxonomien, Dimensionsanalysen und Messinstrumente bilden können. Die Methodik in der taxonomischen Forschung muss weiter standardisiert werden, um einen sprachvergleichenden Forschungsansatz zu ermöglichen und die Konstruktion von Taxonomien in anderen Sprachen zu erleichtern.