Timed picture naming in seven languages

ELIZABETH BATES
University of California, San Diego, La Jolla, California

SIMONA D'AMICO
*University of Rome "La Sapienza," Rome, Italy
and University of Aquila, L'Aquila, Italy*

THOMAS JACOBSEN
University of Leipzig, Leipzig, Germany

ANNA SZÉKELY
Eotvos Lorant University, Budapest, Hungary

ELENA ANDONOVA
New Bulgarian University, Sofia, Bulgaria

ANTONELLA DEVESCOVI
University of Rome "La Sapienza," Rome, Italy

DAN HERRON
University of California, San Diego, La Jolla, California

CHING CHING LU
*National Yang Ming University, Taipei, Taiwan
and National Hsinchu Teachers College, Hsinchu, Taiwan*

THOMAS PECHMANN
University of Leipzig, Leipzig, Germany

CSABA PLÉH
Lorant Eotvos University, Budapest, Hungary

NICOLE WICHA and KARA FEDERMEIER
University of California, San Diego, La Jolla, California

IRINI GERDJIKOVA
New Bulgarian University, Sofia, Bulgaria

GABRIEL GUTIERREZ
University of California, San Diego, La Jolla, California

DAISY HUNG
National Yang Ming University, Taipei, Hsinchu, Taiwan

JEANNE HSU
*National Yang Ming University, Taipei, Taiwan
and National Tsing Hua University, Hsinchu, Taiwan*

GOWRI IYER
University of California, San Diego, La Jolla, California

KATHERINE KOHNERT
*University of California, San Diego, La Jolla, California
and University of Minnesota, Minneapolis, Minnesota*

TEODORA MEHOTCHEVA
New Bulgarian University, Sofia, Bulgaria

ARACELI OROZCO-FIGUEROA
University of California, San Diego, La Jolla, California
and
ANGELA TZENG and OVID TZENG
National Yang Ming University, Taipei, Taiwan

Timed picture naming was compared in seven languages that vary along dimensions known to affect lexical access. Analyses over items focused on factors that determine cross-language universals and cross-language disparities. With regard to universals, number of alternative names had large effects on reaction time within and across languages after target–name agreement was controlled, suggesting inhibitory effects from lexical competitors. For all the languages, word frequency and goodness of depiction had large effects, but objective picture complexity did not. Effects of word structure variables (length, syllable structure, compounding, and initial frication) varied markedly over languages. Strong cross-language correlations were found in naming latencies, frequency, and length. *Other-language* frequency effects were observed (e.g., Chinese frequencies predicting Spanish reaction times) even after within-language effects were controlled (e.g., Spanish frequencies predicting Spanish reaction times). These surprising cross-language correlations challenge widely held assumptions about the lexical locus of length and frequency effects, suggesting instead that they may (at least in part) reflect familiarity and accessibility at a conceptual level that is shared over languages.

Within and across languages, word forms bear little or no resemblance to the concepts that they represent. The same furry four-legged animal is called *dog* in English, *Hund* in German, *perro* in Spanish, and *cane* in Italian. Even within the same language family (e.g., Romance), there are often striking differences in names for the same concept (e.g., *butterfly* is *mariposa* in Spanish, *farfalla* in Italian, and *papillon* in French). Despite these well-known cross-language differences in the shape of words, it is generally assumed that people access their mental lexicon in the same way in every natural language, on the basis of a universal architecture for word comprehension and production. This belief rests crucially on the assumption that the relationship between meaning and form is arbitrary: Word forms have no effect on the process by which speakers move from concept to lexical selection, and meanings have no effect on the shape of words or on the kind of processing required to map a selected concept onto its associated sound.

In the present study, we will investigate some of these fundamental assumptions by presenting what is (to the best of our knowledge) the first large-scale cross-linguistic study of timed picture naming (cf. Bachoud-Levi, Dupoux, Cohen, & Mehler, 1998). We will investigate universal and language-specific contributions to naming behavior across seven languages (English, German, Spanish, Ital-

ian, Bulgarian, Hungarian, and Mandarin Chinese) that vary markedly along lexical and grammatical parameters known or believed to affect word retrieval. This cross-linguistic strategy will shed light on several theoretical questions. Will we find significant differences across languages in *ease of naming* and/or in the particular items that are hard or easy to name? Will such differences reflect cultural variations in the accessibility of concepts and/or linguistic variations in the properties of target names? Will cross-linguistic differences reflect the linguistic distance that separates our seven languages (e.g., Germanic, Romance, and Slavic variants of Indo-European; Uralic; Sino-Tibetan)? Will we find a universal set of predictor–outcome relationships that hold in every language (e.g., effects of frequency on naming latency), and/or will we find language-specific profiles in the relationship between word structure and naming behavior (e.g., effects of syllable structure in Chinese that have no equivalent in English)? Finally, our results will offer insights into the levels at which word and picture properties have their effects. Specifically, for those word properties that are assumed to apply only at the level of word form and/or at the level of lemma selection (Caramazza, 1997; Dell, 1990; Jescheniak & Levelt, 1994), we should find within-language correlations (e.g., English word frequencies predict English reaction times [RTs]) but no corresponding cross-language correlations (e.g., Chinese word frequencies should not predict English RTs). In contrast, if supposed word form effects are confounded by variance from a deeper conceptual level (e.g., word frequency is contaminated by conceptual familiarity/accessibility), we may find significant cross-language effects even after own-language properties are controlled (e.g., Chinese word frequency will predict English RTs, even after English frequencies are controlled).

Timed picture naming was a good candidate for this cross-linguistic strategy for several reasons. Picture naming is a widely used technique for the study of lexical access (for reviews, see Glaser, 1992; Johnson, Paivio, & Clark, 1996), and timed picture naming is one of the first paradigms ever used to study real-time language processing, from early studies by Cattell (1886) through the pioneering work of Wingfield (1967, 1968), Lachman and colleagues (Lachman, 1973; Lachman, Shaffer, & Hennrikus, 1974), and Snodgrass and colleagues (Sanfeliu & Fernandez, 1996; Snodgrass & Vanderwart, 1980; Snodgrass & Yuditsky, 1996). *On-line* (timed) and *off-line* (untimed) picture-naming methods are also widely used in studies of brain-injured patients (for reviews, see Chen & Bates, 1998; Druks & Shallice, 2000; Goodglass, 1993; Murtha, Chertkow, Beauregard, & Evans, 1999) and both normal and language-impaired children (Cycowicz, Friedman, Rothstein, & Snodgrass, 1997; D'Amico, Devescovi, & Bates, 2001; Davidoff & Masterson, 1996; Dockrell, Messer, & George, 2001; Nation, Marshall, & Snowling, 2001). More recently, the neural basis of lexical retrieval has been explored in normal adults, using either overt or covert picture naming, in functional brain imaging (Damasio et al., 2001; Hernandez, Dapretto,

Mazziotta, & Bookheimer, 2001) or event-related brain potentials (Schmitt, Münte, & Kutas, 2000; van Turennout, Hagoort, & Brown, 1997, 1998, 1999; Wicha, Bates, Moreno, & Kutas, 2000).

In our own laboratories, we have used picture naming for several years to explore the time course of lexical access within and across natural languages (Bates et al., 2000). In some studies, pictures have been named out of context, in randomized lists, by monolingual adults (Bates, Burani, D'Amico, & Barca, 2001; Iyer, Saccuman, Bates, & Wulfeck, 2001; Székely & Bates, 2000; Székely et al., 2003; Székely et al., in press), by young children (D'Amico et al., 2001), and by Spanish–English bilinguals across the lifespan (Hernandez et al., 2001; Hernandez, Martinez, & Kohnert, 2000; Kohnert, 2000; Kohnert, Bates, & Hernandez, 1999; Kohnert, Hernandez, & Bates, 1998). In other studies, we have used picture naming to study the effects of phrase and sentence context on lexical access in several languages. Examples include demonstrations of sentence-level semantic priming in English speakers between 3 and 83 years of age (Roe et al., 2000), the effects of grammatical gender on noun retrieval in Spanish (Wicha et al., 2000), Bulgarian (Kokinov & Andonova, unpublished), German (Hillert & Bates, 1996; Jacobsen, 1999), and Italian (Bentrovato, Devescovi, D'Amico, & Bates, 1999), the effects of nominal classifiers on noun retrieval in Chinese (Lu, Hung, Tzeng, & Bates, unpublished) and Swahili (Alcock & Ngorosho, in press), and the differential effects of syntactic cues on retrieval of nouns versus verbs in English (Federmeier & Bates, 1997) and Chinese (Lu et al., 2001).

Although the utility of picture naming is beyond dispute, it has become increasingly clear to us that cross-linguistic comparisons using this technique are limited by the absence of comparable naming norms across the languages of interest. To meet this need, we launched an international project to obtain timed picture-naming norms across a wide range of languages. We have now collected object-naming norms (including both naming and latency) for 520 black-and-white drawings of common objects in seven different languages: American English, Spanish, Italian, German, Bulgarian, Hungarian, and Mandarin Chinese. Initial results from this study are presented below, but first it will be useful to consider in more detail some of the theoretical, as well as the practical, motivations for this work.

Theoretical Basis for
Cross-Linguistic Differences

Why would we expect any systematic differences in picture naming, aside from cultural differences of limited interest for psycholinguistic theories? It is, of course, well known that languages can vary *qualitatively*, in the presence/absence of specific linguistic features that are relevant for lexical access (e.g., Chinese has lexical tone, Hungarian has nominal case markers, and English has neither). In addition, languages can vary *quantitatively*, in the shape and magnitude of the lexical, phonological, and grammatical challenges posed by equivalent struc-

tures for real-time processing and learning. For example, the "same" lexical item (translation equivalents, names for the same pictures) may vary in frequency from one language to another. This might occur for a variety of reasons, including cultural variations in the frequency or accessibility of the concept, as well as cross-language variations in the availability of alternative names for the same concept.

Holding frequency constant, equivalent lexical, phonological, and/or grammatical structures can also vary in their reliability (*cue validity*) and processibility (*cue cost*). These two constructs figure prominently in the competition model (Bates, Devescovi, & Wulfeck, 2001; Bates & MacWhinney, 1989; MacWhinney, 1987), a theoretical framework developed explicitly for cross-linguistic research on acquisition, processing, and aphasia. Within this framework, cue validity refers to the *information value* of a given phonological, lexical, morphological, or syntactic form within a particular language, whereas cue cost refers to the *amount and type of processing* associated with the activation and deployment of that form (e.g., perceivability, salience, neighborhood density vs. structural uniqueness, demands on memory, or demands on speech planning and articulation).

The seven languages that we have selected for study present powerful lexical and grammatical contrasts, with implications for cue validity and cue cost in word retrieval (summarized in Table 1). Hungarian is a Uralic language (from the Finno-Ugric subclass) and is one of the few non-Indo-European languages in Europe. Chinese is a Sino-Tibetan language with strikingly different features from all of the other languages in our study (e.g., lexical tone, a very high degree of compounding, and no inflectional paradigms). The other five languages are Indo-European, although they represent different subclasses: Bulgarian is a Slavic language, Italian and Spanish are Romance languages, German is the prototypical Germanic language, and English shares history and synchronic features with both Germanic and Romance languages. The seven languages vary along a host of grammatical and lexical dimensions that are known or believed to affect real-time word perception and production. These include degree of word order flexibility (e.g., how many different orders of subject, verb, and object are permitted in each language) and the range of contexts in which subjects and/or objects can be omitted. Both these factors influence the predictability of form class (e.g., the probability that the next word will be a noun, adjective, or verb) and, hence, lexical identity. These languages also vary in the availability of morphological cues to the identity of an upcoming word, including nominal classifiers (in Chinese) and prenominal elements that agree in grammatical gender (in Spanish, Italian, German, and Bulgarian) or case (in German and Hungarian). Phonological features that also assist in word retrieval and recognition include lexical tone (in Chinese), vowel harmony (in Hungarian), and stress (in all these languages except Chinese).

Table 1
Characteristics of the Seven Participating Languages

Linguistic Features	English	German	Italian & Spanish	Bulgarian	Hungarian	Chinese
Indo-European	yes	yes	yes	yes	no	no
Language family	Germanic (strong influence of Romance)	Germanic	Romance	Slavic	Uralic (Finno-Ugric)	Sino-Tibetan
Word order variations*	low	medium	high	high	medium	medium
Inflectional morphology	sparse	rich	rich	rich	rich	none
Omission of constituents in free-standing sentences	not permitted	not permitted	subject can be omitted	subject can be omitted	subject can be omitted	subject and object can be omitted
Use of compounding†	high	high	low	medium	medium	high (>80% of all words)
Lexical ambiguity for words out of context	high, especially for nouns and verbs	moderate, especially for nouns and verbs	low for all categories, due to inflectional marking	low for all categories, due to inflectional marking	low for all categories, due to inflectional marking	high for nouns, verbs, and function words
Morphological regularity	one regular and multiple irregular forms for plural and past tense	multiple regular, irregular, and "in-between" (partially productive) forms	multiple regular, irregular, and "in-between" (partially productive) forms	multiple regular, irregular, and "in-between" (partially productive) forms	multiple regular, irregular, and "in-between"(partially productive) forms	lexical regularity only: degrees of productivity in compound formation
Grammatical cues to word identity‡	form class	form class; gender; case	form class; gender	form class; gender	form class; case	form class; nominal classifiers
Prosodic cues to word identity	stress	stress	stress	stress	stress; vowel harmony	lexical tone
Orthography and orthographic regularity	alphabetic; highly opaque/irregular	alphabetic; some irregularities	alphabetic; highly transparent/regular	alphabetic; highly transparent/regular	alphabetic; highly transparent/regular	logographic; one syllable maps to many characters

*Refers to the number of different orders of subject, verb, and object that are possible in the spoken language.   †Refers to words that are composed of other free-standing words (content words and/or function words).   ‡Among grammatical cues to word identity, "form class cues" refer to words or phrases that reliably distinguish between nouns, verbs, and other grammatical classes, as in the difference between "I went to the dance" versus "I want to dance." Studies have shown that such form class cues, like gender, case, and nominal classifiers, can "prime" (facilitate or inhibit) retrieval of words from different grammatical classes.

These seven languages also vary in aspects of word structure that may influence the speed and accuracy of word retrieval outside of a phrase or sentence context. These include large variations in word length, from English (where monosyllabic content words are common) to Italian (where monosyllabic content words are rare). The languages also vary extensively in the use of compounding. In Chinese, more than 80% of all content words are compounds, made up of two or more syllables that occur in many other Chinese words. At the opposite end of the spectrum, compounding is relatively uncommon in Spanish or Italian, with the other languages falling in between. The languages in our data set differ in the amount of lexical homophony that they permit and/or tolerate and in the number of different inflected forms that the lexical target might take. Some of these differences are straightforward and transparent, with documented effects. Others are more subtle, and their effects on lexical access are still unknown.

In the present study, we will demonstrate many similarities across languages in the factors that influence lexical access (e.g., frequency), due in part to the fact that words were accessed out of context, in the simplest possible (*citation*) form. Nevertheless, some intriguing differences have already begun to emerge. Under normal listening conditions, small cross-language differences in average word frequency, length, or complexity may have little or no effect on naming behavior. All languages have evolved under intense communicative pressures, pushing them to "break even" in the overall amount of processing required. Hence, it is likely that any costs associated with (for example) a greater difference in length are compensated for by advantages in other parts of the system. However, the slightly greater demands associated with such cross-linguistic differences may be more evident under adverse processing conditions—for example, in brain-injured patients or in young children. Hence Language A may "break down" before Language B in some situations. Some of the relatively small but significant effects of word structure that we will demonstrate below fall in that category.

Theoretical Basis for Cross-Linguistic Universals

Johnson et al. (1996) have provided a review of picture-naming studies that assumes, on the basis of Paivio's dual-code model (Paivio, 1971), a minimum of three universal stages in word production that are tapped by the picture-naming task: (1) analysis and recognition of the depicted object or event, (2) retrieval of one or more word forms that express the recognized object or event in the speaker's language and selection of the preferred name, and (3) planning and execution of the selected name. No other intervening stages or abstract levels between meaning and sound are assumed. In contrast, Levelt's influential model of word production (which has also relied heavily on picture-naming data; Glaser, 1992; Levelt, 1989; Levelt, Roelofs, & Meyer, 1999) suggests an intervening stage disputed by Johnson et al. Levelt assumes four universal stages in word production: (1) individuation of a target

concept (which could be named in more than one way), (2) selection of a word-specific lemma (the component of a lexical item that contains definitional lexical and grammatical content), (3) activation of the word form (an abstract characterization of the sound pattern associated with a specific lemma, with modifications appropriate for the grammatical context), and (4) articulation of the motor program associated with a word form, modified to fit its articulatory context. Depending on which model ultimately proves to be correct, it is reasonable to assume that natural languages are affected by universal constraints on processing at each of these levels, from conceptualization (perceptual constraints in picture decoding; cognitive and social constraints on concept formation), to retrieval of lexical forms (effects of frequency and complexity), to articulation (universal effects of phonetic structure and length). The search for such universals motivates the selection of measures for multivariate analyses in this cross-linguistic study.

Picture Decoding: Visual Complexity and Goodness of Depiction

As was discussed by Johnson et al. (1996), one element that may contribute to difficulty in picture decoding is visual complexity. In most studies of picture naming that have dealt with this possibility, complexity has been assessed by subjective ratings (e.g., Snodgrass & Yuditsky, 1996). However, Székely and Bates (2000) have shown that subjective ratings of complexity are confounded with subjective judgments of familiarity. To obtain a measure of visual complexity without these confounds, they used the size of the digitized picture file as a predictor variable, measured 10 different ways. These indices proved to be highly correlated with each other and with subjective ratings of complexity but were not related to frequency. One of these measures (see the Materials section) will be used in the present study, on the basis of the assumption that digitized file size should be independent of culture-specific expectations about the "best" way to represent a given concept.

We will also make use of subjective *goodness-of-depiction* ratings by U.S. college students, who were asked to judge how well each picture illustrated the target names that emerged in our study of English (see the Method section). Although it would be preferable to have goodness-of-depiction ratings for all seven languages (a long-term goal), the English ratings did prove to be an excellent predictor of naming behavior across languages, reflecting what may be universal properties of picture decoding.

Conceptual Accessibility

A major goal of this cross-linguistic study was to uncover lexical concepts (as represented in our picture stimuli) that are equally accessible or inaccessible across languages. For this purpose, we have developed indices of *cross-language disparity* and *cross-language similarity*, for both name agreement and latencies (see the Scoring section, below).

Frequency

On the basis of hundreds of studies documenting word frequency effects in lexical access, we expected to find higher name agreement and faster naming latencies for more frequent words (corresponding to word forms in the Johnson et al. model and to lemmas plus their associated word forms in the Levelt model). These frequency effects should be roughly equivalent in magnitude within each of the languages under study. However, the locus of these frequency effects is another matter. As Johnson et al. (1996) noted, conceptual accessibility and word frequency can be hard to distinguish within a given language (is the word frequent because its concept is frequent, or vice versa?). Our cross-linguistic design offers a unique opportunity to learn more about the locus of frequency effects in picture naming. If we find that frequency in Language A predicts name agreement and/or RT not only in Language A, but also in other languages as well, it is likely that conceptual accessibility (in addition to word form frequency) is contributing to the frequency effect (although this would, of course, not rule out frequency effects at other levels as well; see Balota & Chumbley, 1985, for evidence of frequency effects when word naming is delayed well past the point at which the word itself is recognized). Conversely, if frequency in Language A predicts behavior better in Language A than in Language B, it is likely that word form frequency is contributing variance beyond the variance accounted for by the familiarity of the underlying concept. For all seven languages, objective or (in one case) subjective measures of word frequency will be used to compare within- and across-language frequency effects.

Word Length

In previous studies (D'Amico et al., 2001; Székely et al., 2003; Székely et al., in press; see Johnson et al., 1996), modest correlations have been reported between word length and picture-naming behaviors, reflecting lower name agreement and slower latencies for longer names. However, length effects in picture naming appear to be substantially smaller than length effects in word perception tasks (Bates, Burani, et al., 2001; Carr, McCauley, Sperber, & Parmelee, 1982) and are often wiped out in regression analyses removing the effects of confounding variables. One of these confounds is the well-known negative correlation between frequency and word length (i.e., Zipf's law; Zipf, 1965). We may also anticipate confounds between word length and word complexity (see below), as well as conceptual accessibility (more difficult concepts are more likely to be named with longer and more complex words). In our search for universals, we anticipated that word length effects would be small and similar in direction for all the languages, reflecting a universal tendency to avoid longer words. However, we also anticipated cross-linguistic variations in the size of word length effects, which may interact, in turn, with cross-linguistic differences in word structure.

One such example, investigated below, involves the frequency of different *word structure templates*. For example, a large majority of content words in Chinese are disyllables. Each of these syllables is represented by a separate character in the written language and usually has a distinct meaning (which may or may not be modified when that syllable is combined with others to form a compound word). We know from previous studies of Chinese (Chen, 1997; Chen & Bates, 1998) that word structure typicality has effects on naming that are partially independent of word length. That is, disyllables are easier to recognize and/or retrieve than other word types, including monosyllables. In the same vein, monosyllabic content words are common in English but rare in Italian. Hence, any general advantage that may accrue to monosyllables because of their length should be greater in English than it is in Chinese or Italian. To explore the anticipated tradeoff between length and word structure typicality within a cross-linguistic framework, we will include a weighted measure of typicality in the present study, based on the number of target names that fall into each syllable length category (see the Method section).

Phonetic Structure and Word Complexity

For all of the languages under study, we include a dichotomous measure that reflects whether or not the target name produced by our participants begins with a fricative or an affricate (i.e., initial frication). This variable is necessary because it is known that presence of a fricative (which contains white noise, prior to or contemporaneous with voicing) can slow down the detection of word onset. These frication effects are observed in people who participate in timed word perception studies (Bates, Devescovi, Pizzamiglio, D'Amico, & Hernandez, 1995), and they also affect word detection by a computer-controlled voice key in word production studies. Hence, we knew it would be useful to track initial frication in our picture-naming study, even if the relationship between concepts and word forms was completely arbitrary and randomly distributed across natural languages. In addition, we knew in advance that these seven languages share a certain number of cognates, words that are physically similar because they are drawn from a common source. This is true not only for the five Indo-European languages in our study (English, Spanish, Italian, German, and Bulgarian), but also for Hungarian (a Uralic language) and Mandarin Chinese (a Sino-Tibetan language). Hence, the modest nuisance variable *initial frication* may be positively correlated over languages, reflecting a modest degree of physical similarity in the word forms produced within and across these seven languages.

In the same vein, some of the concepts illustrated by our picture stimuli tend to be encoded within and across languages by complex words (compounds or inflected forms) and/or by multiword phrases (e.g., *ice cream cone*). Some items may elicit complex names in every language, reflecting the composite nature of the depicted concept. For example, we anticipated a tendency for all languages to include terms for *fire* and *truck* in the word chosen to name a vehicle driven by people who put out fires. In other cases, complexity effects may be language

specific. For example, Italian tends to use periphrastic constructions like *macchina da stirare* (literally, *machine for ironing*) or *macchina da scrivere* (literally, *machine for writing*), where other languages use a single simple word (e.g., *iron*) or a compound word (e.g., *typewriter*). To deal with universal as well as language-specific effects of word complexity, we included a dichotomous code for word complexity in all multivariate analyses, anticipating that word complexity would be associated with lower name agreement and longer naming latencies. Target names were classified as *complex* if they were compounds or multiword constructions and/or if they contained a marked inflection (e.g., plurals or diminutives). We anticipated that complexity would be confounded with word length, as well as with word frequency, requiring regression designs to assess unique versus overlapping contributions of all three constructs.

The complexity measure was easily obtained for English, Spanish, Italian, German, Bulgarian, and Hungarian, because (1) these languages do have bound inflections and (2) orthographic conventions make it relatively easy to determine whether a word is a compound or a multiword utterance. It was not possible to derive a simple dichotomous measure of complexity for Chinese. First, there is no bound morphology in Chinese (under most definitions). Second, because Chinese writing is logographic rather than orthographic, the distinction between inflected forms, compounds, and/or multiword utterances is controversial. Third, it is estimated that more than 80% of all Chinese content words are compounds; those multisyllabic words that are not decomposable into separate syllables with separate meanings tend to be foreign loan words (e.g., the item in our data set that elicited "sandwich"). Because a dichotomous measure of complexity in Chinese would be difficult to obtain and would, in any case, have an entirely different meaning from the corresponding division in the other six languages under study, Chinese was excluded from all analyses using the complexity measure.

Finally, we discovered in the course of this investigation that speakers sometimes produce the same target name for more than one picture (see also Bates, Burani, et al., 2001; D'Amico et al., 2001; Székely et al., 2003; Székely et al., in press). Although our seven target languages vary in the extent to which this *sharing* strategy is used (see below), it is likely to occur more often for some picture stimuli than others. Specifically, our studies to date suggest that *shared* names are used more often for items that are unfamiliar or difficult to decode. However, the names that are used for this purpose also tend to be shorter and more frequent than those names that emerge as the dominant (*target*) name only once. This phenomenon poses an interesting challenge for multivariate analyses: Because harder items are more likely to share their target names, name sharing may be associated with slower RTs; on the other hand, because shared names tend to be shorter and higher in frequency, name sharing could lead to faster RTs. We will try to disentangle these competing possibilities in regression analyses.

The latter two variables (complexity and name sharing) are examples of historical feedback across the levels postulated by both Paivio's three-stage model and Levelt's four-stage model. Whether or not these relationships between levels are modular within individual speakers/listeners (e.g., the absence of feedback from word form selection to lemma selection is an important constraint in Levelt's theory), the coevolution of concepts and word forms across language history can result in systematic correlations (Kelly, 1992). In other words, even though the relationship between meaning and word form is arbitrary (i.e., words do not resemble their references), word forms can contain correlational cues to meaning, which may affect naming behavior in universal and/or language-specific patterns.

METHOD

Participants

All the participants were native speakers of their respective languages (although amounts of second-language experience may vary with the culture). All were college students, tested individually in a university setting (English speakers in San Diego, Spanish speakers in Tijuana, Mexico, Italians in Rome, Germans in Leipzig, Bulgarians in Sofia, Hungarians in Budapest, and Mandarin Chinese speakers in Taipei). There were 50 participants each in English, Spanish, Italian, Bulgarian, Hungarian, and Chinese; 30 participants were tested in German.

An additional 20 U.S. students participated in a goodness-of-depiction ratings task (see below). Because no adequate objective word frequency norms were available for Bulgarian, a further set of 20 Bulgarian students provided subjective ratings of frequency for the target names obtained with our 520 object pictures, rated on a scale of 1 to 7 from *lowest* to *highest*.

Materials

Picture stimuli for object naming were 520 black-and-white line drawings of common objects. Pictures were obtained from various sources, primarily U.S. and British (listed in Table 2), including 174 pictures from the original Snodgrass and Vanderwart (1980) set. Although different sources were tapped to supplement the Snodgrass and Vanderwart set, all were comparable in style. The full set of picture candidates included more than 1,000 items, many of them overlapping in their content and intended name. Pilot studies (*focus groups*) were carried out in the U.S. for the selection of the final set (e.g., to choose the "best bee" out of three different options). Item selection was subject to several constraints, including picture quality, visual complexity, and potential cross-cultural validity of the depicted item. Nevertheless, the fact that the picture set was originally compiled for English is a limitation of this study and must be kept in mind in comparing results across languages.

The final set of 520 was selected in the hope that each would elicit a distinct object name (although this did not always prove to be the case, even in English). They were scanned and stored digitally for presentation within the PsyScope Experimental Control Shell (Cohen, MacWhinney, Flatt, & Provost, 1993) in 10 different randomized orders. The black-and-white simple line drawings were scanned and saved as (300 × 300 pixels) Macintosh PICT file format, each in a separate file. A demo version of the handmade software utility Image Alchemy 1.8 (Woehrmann, Hessenflow, Kettmann, & Yoshimune, 1994) was used to convert the stimuli to various graphics file formats. Over 30 different file types and degrees of compression for the 520 object and 275 action pictures were computed, and seven commonly used formats were selected according to their relation to subjective visual complexity and other variables. One of these was the Joint Photographic Experts Group (JPEG) (with de-

Table 2
Sources of Object-Naming Stimuli

Source	No.
Snodgrass & Vanderwart, 1980[1]	174
Alterations of Snodgrass & Vanderwart[1]	2
Peabody Picture Vocabulary Test, 1981[2]	62
Alterations of Peabody Picture Vocabulary Test, 1981[2]	8
Martinez–Dronkers set[3]	39
Abbate & La Chappelle "Pictures Please," 1984[4,5]	168
Max Planck Institute for Psycholinguistics[6]	20
Boston Naming Test, 1983[7]	5
Oxford "One Thousand Pictures"[8]	25
Miscellaneous	17

[1] Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, familiarity and visual complexity. *Journal of Experimental Psychology: Human Learning & Memory, 6,* 174-215.

[2] Dunn, Lloyd M., & Dunn, Leota M. (1981). *Peabody Picture Vocabulary Test–Revised.* Circle Pines, MN: American Guidance Service.

[3] Picture set used by Dronkers, N. (personal communication)

[4] Abbate, M. S., & La Chapelle, N. B. (1984b). *Pictures, please! An articulation supplement.* Tucson, AZ: Communication Skill Builders.

[5] Abbate, M. S., & La Chapelle, N. B. (1984a). *Pictures, please! A language supplement.* Tucson, AZ: Communication Skill Builders.

[6] Max Planck Institute for Psycholinguistics, Postbus 310, NL-6500 AH Nijmegen, The Netherlands.

[7] Kaplan, E., Goodglass, H., & Weintraub, S. (1983). *Boston Naming Test.* Philadelphia: Lee & Febiger.

[8] *Oxford Junior Workbooks* (1965). Oxford: Oxford University Press.

fault Huffman coding and high quality–low degree of compression). The syntax used when converting the pictures in Image Alchemy was -j98 (high quality refers to 98 on a scale from 1 to 100). This format was selected for the visual complexity measure (see the Predictor Variables section, below).

Pilot-naming studies indicated that normal adult participants were able to complete 520 items in a single 45- to 60-min session, with occasional breaks. In a paper on English only, Székely et al. (in press) found no significant correlation between accuracy and order of presentation across the naming session. However, naming latencies did slow gradually across the session, justifying our decision to use multiple randomized orders. The same 10 orders were used in every language (3 participants per list in German, 5 participants per list in the remaining six languages).

Procedure

The participants were tested individually in a dimly lit, quiet room. Prior to the picture-naming task, voice sensitivity was calibrated for each participant, who read a list of words with various initial-phoneme patterns (none of these was appropriate as a name for a picture in the main experiment). They were instructed to name the pictures that would appear on the screen as quickly as they could without making a mistake and to avoid coughs, false starts, hesitations (e.g., "uhmm"), articles, or any other extraneous material (e.g., "a dog" or "That's a dog") but to give the best and shortest name they could think of for the depicted object or action. To familiarize the participants with the experiment, a practice set of pictures depicting geometric forms, such as a triangle, a circle, and a square, were given as examples in object naming.

To maximize comparability, identical equipment was used in all seven languages. During testing, the participants wore headphones with a sensitive built-in microphone (adjusted to optimal distance from the participant's mouth) that were connected to the Carnegie Mellon button box, an RT-measuring device with 1-msec resolution design for use with Macintosh computers. The pictures were displayed on a standard VGA computer screen set to 640 × 480 bit-depth resolution (the pictures were 300 × 300 pixels). The participants viewed the items centered, from a distance of approximately 80 cm. On each trial, a fixation crosshatch "+" appeared centered on the screen for 200 msec, followed by a 500-msec blank interval. The target picture remained on the screen for a maximum of 3 sec (3,000 msec). The picture disappeared from the screen as soon as a vocal response was registered by the voice key (at the same time, a dot "•" signaled voice detection—a clue for the error-coding procedure). If there was no response, the picture disappeared after 3,000 msec, but another 1,000 msec were added to the total response window, just in case the speakers had initiated a response right before the picture disappeared. Hence, the total window within which a response could be made was 4,000 msec. The period between offset of one trial and onset of the next was set to vary randomly between 1,000 and 2,000 msec. This intertrial *jitter* served to prevent subjects from settling into a response rhythm that was independent of item difficulty.

RTs were recorded automatically by the voice key in the CMU button box and served as critical outcome measures for statistical analysis. For each of the 10 different randomized versions of the experiment, a printout served as a score sheet for coding purposes during the experiment. The experimenter took notes on the score sheet according to an error-coding protocol (see details below). Alternative namings were also recorded manually on the score sheet. No pictures were preexposed or repeated during the test; hence, no training of the actual targets occurred. A short rest period was included automatically after 104 trials, but the participants could ask for a pause in the experiment at any time. Experimental sessions lasted 45 min on average and were tape-recorded for subsequent off-line checking of the records.

To obtain subjective ratings of goodness of depiction, the same 520 object pictures were presented to a separate group of 50 U.S. college students. The students were asked to rate how well the picture fit its dominant name (determined empirically from the English picture-naming study; see the Scoring section) on a 7-point scale from *worst* to *best*. A keypress recording procedure allowed the participants to use a broad time interval for making their responses, since the stimulus remained on the screen until the participant responded by pressing one of the keys representing the scales. Average ratings were calculated for each picture.

Scoring

Our scoring criteria were modeled closely on procedures adopted by Snodgrass and Vanderwart (1980), with a few exceptions. The target name (dominant response) for each picture was determined empirically in two steps.

First, error coding was conducted to determine which responses could be retained for both naming and RT analyses. The following three error codes were possible.

1. *Valid response* referred to all the responses with a valid (codable) name and valid response times (no coughs, hesitations, false starts, or prenominal verbalization, such as "that's a ball"). Any word articulated completely and correctly was kept for the evaluation, except for expressions that were not intended namings of the presented object, such as "I don't know."

2. *Invalid response* referred to all the responses with an invalid RT (i.e., coughs, hesitations, false starts, or prenominal verbalizations) or a missing RT (the participant did produce a name, but it failed to register with the voice key).

3. *No response* referred to any trial in which the participant made no verbal response of any kind.

Once the set of valid responses had been determined, the target name was defined as the *dominant response*—that is, the name that was used by the largest number of participants. In the case of ties (two responses uttered by exactly the same number of participants) three criteria were used to choose one of the two or more tied responses as the target: (1) the response closest to the intended target (i.e., the hypothesized target name used to select stimuli prior to the experiment),

(2) the singular form, if singular and plural forms were tied, or (3) the form that had the largest number of phonological variants in common.

Second, all valid responses were coded into the following different lexical categories in relation to the target name, using the same criteria.

Lexical Code 1. The target name (dominant response, empirically derived).

Lexical Code 2. Any morphological or morphophonological alteration of the target name, defined as a variation that shares the word root or a key portion of the word without changing the word's core meaning. Examples would include diminutives (e.g., *bike* for *bicycle*, *doggie* for *dog*), plural/singular alternations (e.g., *cookies* when the target word was *cookie*), reductions (e.g., *thread* for *spool of thread*), or expansions (e.g., *truck for firemen* for *firetruck*).

Lexical Code 3. Synonyms for the target name (which differ from Code 2 because they do not share the word root or key portion of the target word). With this constraint, a synonym was defined as a word that shared the same truth-value conditions as the target name (e.g., *couch* for *sofa* or *chicken* for *hen*).

Lexical Code 4. This category was used for all names that could not be classified in Codes 1–3, including hyponyms (e.g., *animal* for *dog*), semantic associates that are not synonyms (e.g., *cat* for *dog*), part–whole relations at the visual-semantic level (e.g., *finger* for *hand*), and all frank visual errors or completely unrelated responses.

Name Agreement

Percentage of name agreement was defined, for each item, as the proportion of all valid trials (a codable response, with a usable RT) on which the participants produced the target name (*Lex1*). The number of alternative names for each picture (*number of types*) was derived by simply counting number of different names provided on valid trials, including the target name. In addition, following Snodgrass and Vanderwart (1980), we also calculated the *H* statistic, or *H* Stat (also called the *U* statistic), a measure of response agreement that takes into consideration the proportion of participants producing each alternative. High *H* values indicate low name agreement, and 0 refers to perfect name agreement. Percentage of name agreement measures for each item were based on the 4-point lexical coding scheme. For each item, *Lex1* refers to the percentage of all codable responses with a valid RT on which the dominant name was produced.

Although the 520 object pictures were selected with the intention of eliciting one unique target name for every picture, this was not always the case. Occasionally, within a given language, the same name emerged as the dominant response for two or more pictures. We treated this event as a dependent variable (called *sames*), assigning a score of 1 for each item that shared its target name (dominant response) with another picture in the data set for that language. This assignment was made independently for each language. Hence, an item might have a unique name in English (esames = 0) but share its name in Bulgarian (bsames = 1), and so forth.

Reaction Time

RT total refers to mean RTs across all valid trials, regardless of the content of that response. *RT target* refers to mean latency for dominant responses only.

Cross-Language Universality and Disparity

The cross-linguistic design of the present study permitted us to derive some novel measures of *universality* and *cross-language disparity* for the 520 picture stimuli. For this purpose, name agreement (Lex1) and latencies to produce the target name were first converted to *z* scores within each language (representing a continuum from *best* to *worst* in name agreement and from *fastest* to *slowest* in target RT). By using *z* scores, we removed main effects of language on both of these dependent variables and ensured that no single lan-guage was contributing disproportionately to these estimates of universality and disparity. Estimates of universality were obtained by averaging the *z* scores for each item, across all seven languages, for Lex1 and RT target, respectively. Thus, if an item tended to elicit high agreement and fast RTs in all or most of the languages, it would have an average universal *z* score at the high-performance end of each continuum (one for name agreement, one for RT). Conversely, if an item tended to elicit low agreement and/or slow RTs in all or most of the languages, it would have an average universal *z* score at the low-performance end of each continuum. Items that produced little consensus across the seven languages would tend to cluster in the middle of these two universality measures. A third, relatively simple measure of universal tendencies was also constructed, for each item, on the basis of the arithmetic average of the *number of word types* measure for each of the seven languages.

To complement these measures of universality, we also calculated measures of cross-language disparity. We began by computing, for each language, the average *z* scores for the other six languages (for Lex1 and for RT target, respectively). So, for example, the *Other-Language Z-score* for English name agreement would be the average of the *z* scores for German, Spanish, Italian, Bulgarian, Hungarian and Chinese. In the same vein, the *Other-Language Z-score* for Bulgarian RT target would be the average of the RT *z* scores for English, German, Spanish, Italian, Hungarian, and Chinese. With these statistics in hand, we then calculated a difference score for each language, for agreement and RT respectively, using simple subtraction in a direction that would indicate an advantage for the language in question. For example, a positive Lex1 difference score for German would indicate that Germans had higher name agreement for that item than did the other six languages (on average). Similarly, a negative RT target difference score for Chinese would indicate that Chinese speakers were relatively fast on that item, as compared with the speakers in the other six languages (on average). Using these difference scores, we can investigate the factors that confer a relative advantage (*easy item*) or disadvantage (*hard item*) within each individual language, as compared with the others in the study. Finally, the absolute values of these seven difference scores were averaged, for both Lex1 and RT target, to produce estimates of cross-language disparity in nameability (Lex1) and RT (RT target), respectively. Thus, items with high disparity scores are those that elicited more cross-language variation; items with low disparity scores are those that elicited less variability and a more universal response (although such items could be *universally good* or *universally bad*).

In addition to these measures of performance (dependent variables), the target (dominant) names produced for each item were coded along a number of dimensions that were believed to affect accuracy and/or latency in studies of lexical access (independent variables). In the full database, this list includes some variables that are applicable (e.g., grammatical gender) or available (e.g., age-of-acquisition [AoA] norms) only for a subset of the languages. For the present study, we restricted our attention to the following independent or predictor variables that were available for all seven languages (with two exceptions, indicated below).

1. *Visual complexity*. In addition to predictor variables associated with the target names, an estimate of visual complexity was obtained for each picture, based on file size in the JPEG format (see the Materials section, above; for additional details, see Székely & Bates, 2000).

2. *Goodness-of-depiction ratings* were available only for U.S. participants, but this variable proved to be such a powerful predictor across all seven languages that it was included in the present study.

3. *Word frequency* of the target names was extracted for each language from written or spoken sources (because there were no frequency corpora available for Bulgarian, subjective ratings of frequency were used). The sources included the following: the CELEX database for English and German (Baayen, Piepenbrock, &

Gulikers, 1995); Alameda and Cuetos (1995) for Spanish; De Mauro, Mancini, Vedovelli, and Voghera (1993) for Italian; Füredi and Kelemen (1989) for Hungarian; and the Chinese Knowledge Information Processing Group (1997) for Chinese.

4. *Word length in syllables*.

5. *Syllable type frequency*. As we noted in the introduction, the seven languages vary markedly in the distribution of monosyllables, disyllables, and words with three or more syllables. After the target names were ascertained for each language and their length in syllables was calculated, we constructed a measure to reflect the frequency of word types in the full corpus of 520 names for each language. For example, if the corpus for Language A comprised 220 monosyllables, 200 disyllables, 80 three-syllable words, 19 four-syllable words, and 1 five-syllable word, each monosyllabic name received a score of 220, each disyllable received a score of 200, each three-syllable word received a score of 80, each four-syllable target name received a score of 19, and the remaining five-syllable word received a score of 1.

6. *Word length in characters* was available for all the languages except Chinese.

7. *Initial frication* was a dichotomous variable reflecting presence/absence of a fricative or affricate in the initial consonant (0 = no fricative or affricate, 1 = fricative or affricate).

8. *Complex word structure* was a dichotomous variable that was assigned to any item on which the dominant response was an inflected form, a compound word, or a periphrastic (multiword) construction. This was available for all languages except Chinese.

RESULTS AND DISCUSSION

All the analyses were conducted over items, averaged across the 50 subjects (30 for German) who participated in the naming study for each language. Hence, language was treated as a within-item factor for all direct cross-linguistic comparisons. Within all the relevant tables, languages are ordered to reflect hypothesized differences in *language distance*, from English and German (Germanic), to Spanish and Italian (Romance), to Bulgarian (Slavic), and then to the two non-Indo-European languages, Hungarian and Chinese. Because these results are unusually complex (involving seven languages, multiple dependent variables, and multivariate analyses), we have tried to make the text more readable by placing statistical details into tables wherever possible. Hence, the text itself will focus on overall patterns that emerged

from these detailed results. Some of the tables are referred to in the text (in appropriate order), but for the sake of economy, they are excluded from the published text and are, instead, available for inspection (together with those pictures that are publicly available) on our Web site at http://www.crl.ucsd.edu/~aszekely/ipnp/7lgpno.html.

Part I: Cross-Language Variations in Naming Behavior (Dependent Variables)

Part I presents cross-linguistic similarities and differences in naming behavior (dependent variables), including various measures of name agreement, as well as latencies to produce the target name. This section also includes multivariate analyses of the relationships among these dependent variables, within and across languages. This is the place where we can ask our first three questions about cross-language disparities and cross-language universals in naming behavior.

Question I-1. To what extent will naming behaviors vary across these seven languages? If differences in naming behavior are detected, will they reflect a classic gradient of language distance (e.g., stronger correlations among Indo-European languages vs. Hungarian and Chinese; stronger correlations within specific Indo-European language families, such as Romance).

Question I-2. Will cross-language correlations be similar for name agreement and reaction times, reflecting *basic conceptual universals*? Or will we find stronger cross-language similarities in the time required to retrieve the dominant target name, reflecting universal aspects of processing that cannot be detected with name agreement alone?

Question I-3. Will the number of alternative names for each picture slow down RTs (reflecting competitor effects), even after percentage of agreement on the dominant name is controlled? If such competitor effects can be demonstrated, will they qualify as a cross-language universal, evident within every language and in cross-language summary scores?

Descriptive statistics. A breakdown of the proportion of valid responses, invalid responses, and nonresponses observed within each language (means, standard devia-

Table 3
Summary Statistics for Correctness in the Different Languages

	English	German	Spanish	Italian	Bulgarian	Hungarian	Chinese
% valid response ($F = 89.56, p < .001$)							
Mean	96.1	94.7	93.2	92.0	89.2	94.1	89.3
SD	6.0	9.6	10.3	10.9	11.1	8.2	11.9
Range	60–100	17–100	34–100	18–100	20–100	22–100	22–100
% no response ($F = 35.60, p < .001$)							
Mean	2.3	3.3	5.2	5.5	5.1	2.2	4.6
SD	5.0	8.8	9.6	9.9	10.0	6.7	10.1
Range	0–34	0–80	0–66	0–80	0–78	0–74	0–76
% invalid response ($F = 191.64, p < .001$)							
Mean	1.5	2.0	1.6	2.5	5.7	3.7	6.1
SD	2.3	3.0	2.1	2.8	4.5	3.3	4.2
Range	0–16	0–20	0–14	0–14	0–32	0–20	0–22

tions, and ranges) is presented in Table 3. One-way analyses of variance over items revealed significant cross-linguistic differences on all three variables. In part, this reflected an overall advantage for English (with a mean of 96.1% valid responses with usable RTs) and a relative disadvantage for Bulgarian and Chinese (at 89.2% and 89.3%, respectively). The fact that performance was better in English is not at all surprising, since the items were initially developed and selected for an English-speaking audience. However, this is not the whole story: When the same analyses of variance were repeated with English excluded, significant cross-language differences remained for the other six languages on all three measures in Table 3 (with F values between 31.1 and 176.7, all ps < .001). Although there is certainly cross-language variation on all three measures, the most important finding in Table 3 is that this naming paradigm works in all seven languages and also yields enough variation in *nameability* within each language to justify the multivariate approach that we take for the remainder of the paper.

Descriptive statistics for more detailed name agreement variables within each language are presented in Table 4. There were robust cross-linguistic differences on every measure (assessed by a one-way analysis of variance for all continuous variables and by a k-level chi-square for the binary variable *same name*). Performance was highest for English speakers in every case (as we would expect, since the items were originally compiled from English sources). However, significant cross-language differences remained when the same analyses were repeated excluding English (all ps < .001). Chinese speakers were almost always at a greater disadvantage (lower name agreement and more alternative names). But target name agreement was high in every language group, from a low of 71.9% for Chinese to a high of 85.0% in English.

Naming latencies for each language are presented in Table 5 for all valid trials (total RT) and for those trials on which speakers produced the target name (target RT). One-way analyses of variance over languages again revealed significant differences, with the fastest RTs observed in English ($M = 1,041$ msec) and the slowest RTs observed in Bulgarian ($M = 1,254$ msec) and Chinese ($M = 1,241$ msec). However, the range of RTs was very large within every language, and the fastest RTs were quite comparable (ranging from 656 msec in English to 768 msec in Bulgarian). Both the mean RTs and the minimum RTs are comparable to values that have been reported in prior studies using timed picture naming.

Table 4
Summary Statistics for Name Agreement in the Different Languages

	English	German	Spanish	Italian	Bulgarian	Hungarian	Chinese
Number of types ($F = 58.43, p < .001$)							
Mean	3.35	5.14*	4.15	4.39	3.82	4.16	5.47
SD	2.28	3.42*	2.91	2.85	2.56	2.96	3.63
Range	1–18	1.7–21.7*	1–17	1–20	1–14	1–21	1–21
H statistics ($F = 46.65, p < .001$)							
Mean	0.67	0.76	0.86	0.95	0.84	0.91	1.16
SD	0.61	0.68	0.72	0.73	0.65	0.73	0.79
Range	0–2.90	0–3.28	0–2.90	0–3.47	0–2.70	0–3.52	0–3.57
% Lex 1 dominant ($F = 32.83, p < .001$)							
Mean	85.0	81.1	80.0	77.0	80.2	78.0	71.9
SD	16.4	19.9	20.4	21.6	20.4	21.3	23.3
Range	28–100	21–100	17–100	12–100	13–100	13–100	11–100
% Lex 2 phonetic variance ($F = 21.64, p < .001$)							
Mean	3.7	4.4	3.2	4.9	4.1	7.1	8.5
SD	8.7	10.0	8.4	10.4	9.8	12.9	12.4
Range	28–100	21–100	17–100	12–100	13–100	13–100	11–100
% Lex 3 synonym ($F = 11.78, p < .001$)							
Mean	2.4	3.2	4.2	5.2	2.5	4.3	1.6
SD	7.7	8.4	10.1	11.0	7.7	10.2	5.5
Range	28–100	21–100	17–100	12–100	13–100	13–100	11–100
% Lex 4 erroneous ($F = 29.19, p < .001$)							
Mean	9.0	11.4	12.7	12.9	3.3	10.6	18.0
SD	12.4	16.4	16.2	16.4	17.4	16.2	19.8
Range	28–100	21–100	17–100	12–100	13–100	13–100	11–100
Same name ($\chi^2 = 91.2, p < .001$)							
Mean (%)	4.6	8.3	12.1	8.7	12.9	14.0	19.6
Range	0–1	0–1	0–1	0–1	0–1	0–1	0–1

*Since data were collected from only 30 subjects in the German language, the number of alternative types were calculated as (raw type number) × (50/30).

Table 5
Summary Statistics for Mean Reaction Time in the Different Languages

	English	German	Spanish	Italian	Bulgarian	Hungarian	Chinese
RT total ($F = 136.76, p < .001$)							
Mean	1,041	1,130	1,168	1,163	1,254	1,105	1,241
SD	230	281	280	270	283	281	319
Range	656–1,843	663–2,397	711–2,063	694–2,580	768–2,373	659–2,300	686–2,389
RT target ($F = 115.00, p < .001$)							
Mean	1,019	1,101	1,139	1,133	1,217	1,071	1,200
SD	211	273	262	264	261	268	312
Range	656–1,823	663–3,117	711–2,392	694–2,831	768–2,273	659–3,139	686–2,403

Again, when the cross-language analyses were conducted with English excluded, significant main effects of language remained.

Correlations and regressions within languages for dependent variables. Tables 6–12 (on our Web site at http://www.crl.ucsd.edu/~aszekely/ipnp/7lgpno.html) present correlations among eight dependent measures, within each language, including a summary score for *percent correct* that conflates Lexical Codes 1, 2, and 3 (targets + morphophonological variants + synonyms). This "lenient" score for name agreement has been used in other picture-naming studies and sometimes yields different results, as compared with the more conservative measure in which name agreement refers to production of the target name only (see Székely et al., in press, for details). In the present study, the results for these two versions of name agreement were quite similar; hence, we will restrict our attention to the strict (Lex1) measure of name agreement for most of the analyses that follow. Most important for present purposes, inspection of Tables 6–12 indicates that there are striking similarities across languages in the pattern of correlations. Of course, some of these correlations are inevitable, because variables such as percentage of target names and the *H* statistic stand in a part–whole relationship. Others are more interesting, including high negative correlations between target name agreement (Lex1) and target RT. The least obvious statistics in Tables 6–12 involve small but significant correlations with the binary variable *same name*. In general, these results suggest that speakers tend to produce the same name for more than one picture when they find themselves in difficulty, a strategy that is least likely in English and most likely in Chinese but may re-

flect similar processes in all seven languages (for further discussion of this point, see Székely et al., 2003).

To complement the language-specific statistics in Tables 6–12, Table 13 presents correlations among the five summary measures of universality and disparity, averaged across the seven languages. The cross-language pattern is similar to the relationships seen within individual languages. However, the coefficients for these summary variables are notably higher than the corresponding coefficients for each individual language. For example, the correlation between target agreement and target RT ranges between −.55 and −.67 in the seven individual languages, but the corresponding relationship between universal agreement and universal RT is −.74. Similarly, the correlations between target RT and number of alternative names range between +.61 and +.76 within individual languages, but the cross-language relationship between target RT and average number of types is +.81. This result is not trivial. In summarizing over seven different language types, we might have lost information based on specific language forms (e.g., item-specific effects of frequency in German or length in Hungarian) that contributes to the magnitude of the correlations within each language. *If such form-specific facts played a major role in the correlations among naming measures, we would expect these coefficients to drop markedly when cross-language summary variables are used.* Instead, we seem to be picking up power and reliability by averaging across languages, reflecting strong universal trends that are *greater than the sum of the parts.*

Finally, Table 13 shows that cross-language universals are inversely correlated with corresponding measures of cross-language disparity. Although this is not surprising

Table 13
Correlation Matrix of Cross-Language Dependent Variables

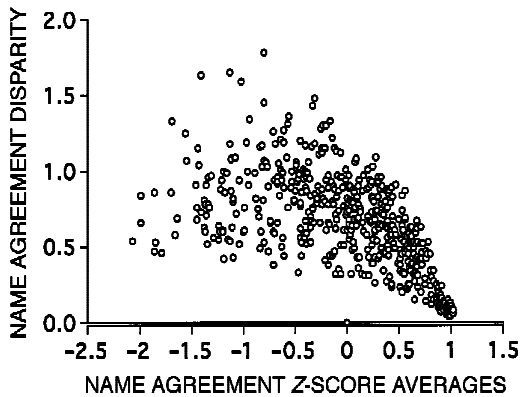|	Cross-Language Name Agreement	Cross-Language Target RT	Cross-Language Number of Types	Cross-Language Naming Disparity
Cross-language name agreement	–			
Cross-language target RT	−.74‡	–		
Cross-language number of types	−.89‡	+.81‡	–	
Cross-language naming disparity	−.60‡	+.40‡	+.50‡	–
Cross-language target RT disparity	−.53‡	+.66‡	+.55‡	−.51‡

‡$p < .01$.

Figure 1. Scatterplot of name agreement disparity scores plotted against name agreement *z* scores (both averaged over languages).

(indeed, it was inevitable), it is interesting that the relationships are lower than we might expect. For example, the correlation between universal name agreement and cross-language naming disparity is −.60. This means that items high in disparity tend to be low in agreement, but it also means that the two measures share only 36% of their variance. To explore the relationship between universality and disparity further, we examined scatterplots for two pairs of variables: universal name agreement versus naming disparity (Figure 1) and universal target RTs versus RT disparity (Figure 2). Both of these scatterplots show that *good* items are very good indeed (high in agreement, low in RT, and low in disparity on both measures). However, *bad* items can be universally bad (low agreement, large RTs, and low disparity on both measures) or they can elicit quite diverse results (bad for only a subset of languages, with high disparity).

The raw correlations in Tables 6–13 suggest another question of considerable theoretical importance: Is *nameability* a single dimension that bears a singular relationship to naming latency, or can we distinguish between nameability (reflected in proportion of participants producing the dominant name) and competition from alternative names? To deal with this question, we conducted stepwise regression analyses for target RT in each language, using name agreement (Lexical Code 1) and number of types as predictors. Each regression was conducted twice, giving each predictor an opportunity to enter the equation last, in order to assess the amount of unique variance contributed by that measure when the other was controlled. We used number of types instead of the *H* statistic for this purpose, because the *H* statistic (which is weighted for the number of participants choosing each alternative) and Lexical Code 1 are correlated so highly that they are likely to cancel each other out.

Table 14 summarizes the results for these stepwise analyses within each language, and it also presents the corresponding results when cross-language summary variables are used. In all seven languages, the two naming

measures together accounted for a substantial amount of the variance in naming latencies (from a low of 39.2% in German to a high of 58% in Spanish). In all seven cases, number of word types contributed significant and substantial variance to naming latencies after name agreement was controlled (from a low of 6.0% in German to a high of 17.1% in English). In all cases, the direction of that contribution was positive, indicating that RTs slowed down as a function of the number of alternative names that were available, independent of name agreement for the dominant response. This result is compatible with models in which alternative names exert a competitive effect on word retrieval (Caramazza, 1997; Cutting & Ferreira, 1999; Humphreys, Riddoch, & Quinlan, 1988; Levelt et al., 1999; Roelofs, 1992; Schriefers, Meyer, & Levelt, 1990). In all seven languages, target name agreement also made a significant contribution after number of word types was controlled, always in a negative direction (indicating that higher agreement is associated with faster RTs). However, these contributions were considerably smaller (from a barely significant low of −0.5% in English to a high of −4.4% in Italian). Hence, we may conclude that name agreement and alternative names exert at least partially independent effects on the picture-naming process but that the inhibitory effect of alternative names is larger than the facilitative effect of name agreement.

These appear to be universal effects, although they vary in magnitude from one language to another. In this regard, the regression using cross-language summary variables is particularly informative. Using universal RT as the dependent variable (i.e., the cross-language average for *z* score target-naming latencies), the cross-language averages for number of types and for name agreement account together for a significant 66.1% of the RT variance (*p* < .001). This statistic is (once again) larger than the corresponding figures obtained within individual languages, indicating that we have gained in power and reliability by summarizing across languages. The cross-language average for number of word types made a large
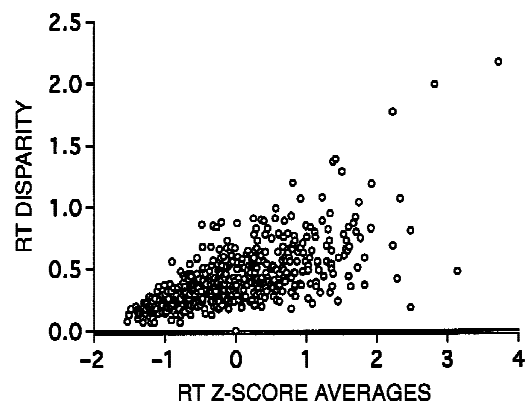

Figure 2. Scatterplot of reaction time (RT) disparity scores plotted against RT *z* scores (both averaged over languages).

Table 14
Regressions of Naming Behavior on Naming Latencies Within Each Language and Across
Languages (Cross-Language Average Z Score Reaction Times [RTs]): Total Percentage of
Variance Accounted for and Unique Contributions of Each Variable on the Last Step

Languages	% Total Variance	% Unique Variance From Name Agreement	% Unique Variance From No. of Types
English	47.3‡	−0.5†	+17.1‡
German	39.2‡	−2.1‡	+ 6.0‡
Spanish	58.0‡	−0.7‡	+13.6‡
Italian	46.0‡	−4.4‡	+5.9‡
Bulgarian	52.5‡	−1.7‡	+11.6‡
Hungarian	51.0‡	−1.0‡	+11.9‡
Chinese	48.9‡	−1.9‡	+8.7‡
Cross-language Z score RT average	66.1‡	n.s.	+12.1‡

†$p < .05$.   ‡$p < .01$.

and significant contribution (+12.1%, $p < .01$) to universal naming latencies when universal name agreement was controlled, at a level midway between the unique contributions observed in each individual language. However, in contrast with the analyses conducted within each individual language (in which name agreement always made a small but significant independent contribution), the summary score for name agreement had no significant effect when the summary score for number of types was entered into the equation first. *In other words, the most important determinant of RTs within and across languages appears to be the number of alternative names that are available, rather than the level of agreement that was reached for the target name itself.*

Correlations and regressions across languages for dependent variables. Tables 15–18 (on our Web site at http://www.crl.ucsd.edu/~aszekely/ipnp/7lgpno.html) present cross-language correlations for name agreement, number of alternative names, same names, and target name agreement.

For name agreement, all correlations are significant ($p < .01$), suggesting that there was some cross-language generalization in the relative difficulty (nameability) of these 520 picture stimuli. We might have expected the magnitude of these correlations to follow a typological metric of language distance, but there was relatively little evidence for such a gradient. Indeed, the highest correlations often involve Hungarian (which is not Indo-European) with other European languages. Correlations with Chinese do tend to be lower, but not enough to suggest that this language is an outlier.

Table 16 summarizes correlations across languages for number of alternative names. Coefficients for number of word types are in the lower triangle; coefficients for the H statistic are in the upper triangle. Again, all of these correlations are robust ($p < .01$) and moderate in size. In other words, there is cross-language generality in the tendency for some items to elicit multiple names.

Table 17 presents correlations across languages for the dichotomous variable *same name*. Although these coefficients are smaller than the other naming variables, they are all highly significant ($p < .01$) and all in the

same direction. This result provides support for our hypothesis that speakers tend to share *generic* names for difficult items—which are, in many cases, the same items from one language to another.

Finally, cross-language correlations for naming latencies are presented in Table 18. All these correlations are significant ($p < .01$), and they are also very large. The magnitude of these correlations is striking, in comparison with the more modest coefficients observed for various measures of name agreement (Tables 15–17). However, there is no obvious typological gradient underlying these correlations, and no evidence that any individual language is an outlier. Instead, these results suggest that *naming latencies are influenced by universal stages and processes, shared by the widely varying languages in our sample.* These shared processes are more evident in the time required to produce a name than they are in the extent to which speakers agree on the names that they produce.

Part I: Interim Summary

To summarize the results so far, let us return to the three main questions posed in Part I.

Question I-1. To what extent will naming behaviors vary across these seven languages? If differences in naming behavior are detected, will they reflect a classic gradient of language distance (e.g., stronger correlations among Indo-European languages vs. Hungarian and Chinese; stronger correlations within specific Indo-European language families, such as Romance)?

The seven languages under study here did differ significantly on all indices of name agreement and in the time required to retrieve and produce those names. English had the advantage on all the measures that we used—not a surprising result, since these pictures were designed for use in American or British English studies and were compiled for the present study in an English-speaking context. However, all cross-language comparisons remained robustly significant when English was excluded from the analysis. The lowest scores for name agreement and the longest RTs were typically observed in Chinese and/or Bulgarian. However, the cross-language differences in performance observed so far provide little

evidence for a typological gradient. That is, there is little evidence here for greater similarities in performance within language families or for greater disparities between language families.

Question I-2. Will cross-language correlations be similar for name agreement and RTs, reflecting *basic conceptual universals*? Or will we find stronger cross-language similarities in the time required to retrieve the dominant target name, reflecting universal aspects of processing that cannot be detected with name agreement alone?

There were moderate to strong cross-language correlations for all of the dependent variables, suggesting that at least some of these 520 items tend to be *hard* or *easy* in every language. But correlations were substantially larger in the RT data, suggesting that this timed picture-naming paradigm taps into universal stages and/or universal processes that transcend *nameability* itself. These results were even stronger for our cross-language summary variable, as compared with correlations within each individual language. This is, we must reiterate, a nontrivial result. If language-specific differences in word properties were playing a major role in our results, within-language correlations among naming behaviors ought to be stronger than across-language correlations. The fact that our novel cross-linguistic dependent variables worked so well suggests that we gained in power and reliability by pooling results over languages.

Question I-3. Will the number of alternative names for each picture slow down RTs (reflecting competitor effects), even after percentage of agreement on the dominant name is controlled? If such competitor effects can be demonstrated, will they qualify as a cross-language universal, evident within every language and in cross-language summary scores?

In all seven languages, the number of alternative names for each picture contributed substantial variance to RTs after name agreement was controlled. When cross-language summary scores were used, results were even stronger. This universal result is compatible with models that emphasize competitor effects during name retrieval.

We turn now to cross-language similarities and differences in the properties of those words that emerged empirically as target names (also called *dominant response*, or Lexical Code 1), in relation to each other and to properties of the picture stimuli.

Part II: Word and Picture Properties (Independent Variables)

Part II presents cross-linguistic findings for the stimulus properties (independent variables) that characterize the target pictures (visual complexity and goodness of depiction) and the target names elicited in each language (e.g., frequency, length, and word structure). Multivariate analyses will focus on similarities and differences in the pattern of relationships among independent variables within and across languages. In this section, we can address the following four questions about cross-language universals and cross-language disparities.

Question II-1. How much will languages differ in the properties of target names that are elicited by the same 520 pictures of common objects?

Question II-2. Will picture properties (objective complexity and goodness of depiction) be independent of each other, and will they be independent of the properties of target names within each language? If we do find picture–name confounds (e.g., complex pictures elicit more complex names), will these be language specific, or will they generalize across languages in the study?

Question II-3. Will we replicate Zipf's law across all the languages in the study? That is, will frequency be associated universally with longer words? And will the same law extend to other properties of word structure (e.g., canonical syllable structure)?

Question II-4. Will we find correlations in word characteristics across the seven languages in this study? If the link between meaning and word form is truly arbitrary, we would not expect to find cross-language correlations in the properties of target picture names (e.g., frequency, length, or word structure). But if, instead, we do find strong cross-language correlations in word properties (e.g., word frequency in Language A correlates with word frequency in Language B), a case can be made that these correlations reflect (at least in part) universals at a deeper level in the picture-naming process (e.g., conceptual accessibility).

Descriptive statistics. Table 19 provides descriptive statistics for each language for five properties of the target names that emerged for each picture stimulus in that language: length in syllables, length in characters (excluding Chinese), word frequency (based on a different metric for every language, including subjective frequency ratings in Bulgarian vs. log natural frequency from available corpora for the other six), initial frication (a binary variable), and word complexity (also a binary variable, not available for Chinese). One-way within-item analyses of variance were conducted on the two length measures; chi-square (*k*-level) analyses were conducted on the two binary variables. The seven frequency indices were not compared directly, because they were based on very different counts.

As Table 19 shows, there were robust cross-linguistic differences in every relevant category, reflecting differences in word structure of the sort that were described in the introduction (see summary in Table 1). The shortest words were produced in English (*M* = 1.74 syllables; range, 1–5) and the longest words were produced in Italian (*M* = 2.92 syllables; range, 1–8), with the other languages distributed in between at roughly even intervals. There were also marked variations in the probability that target names will begin with a fricative or affricate, from a low of 12.3% in Spanish to a high of 34.8% in Hungarian, a factor that could skew naming latencies detected by a voice key. There were substantial differences in word complexity as well, ranging from 8.1% of the

Table 19
Summary Statistics for Independent Variables Within Each Language

	English	German	Spanish	Italian	Hungarian	Bulgarian	Chinese
Length in syllables ($F = 169.50, p < .001$)							
Mean	1.74	2.13	2.76	2.92	2.28	2.40	2.09
SD	0.83	0.87	0.96	1.00	0.97	0.92	0.60
Range	1–5	1–6	1–7	1–8	1–8	1–7	1–5
Length in characters ($F = 29.08, p < .001$)							
Mean	5.89	6.73	6.48	7.07	6.07	6.29	not relevant
SD	2.22	2.74	2.14	2.50	2.28	2.23	
Range	2–15	2–19	3–17	2–20	2–19	3–17	
Word frequency (no statistical comparison)							
Mean	2.50	2.01	2.90	1.16	1.38	4.25*	3.36
SD	1.57	1.50	1.73	1.43	1.93	1.09*	2.01
Range	0–7.40	0–6.62	0–8.32	0–6.20	0–6.84	1.5–6.8*	0–10.56
Frication (%) ($\chi^2 = 85.34, p < .001$)							
Mean	28.1	27.5	12.3	24.8	34.8	27.1	25.8
Range	0–1	0–1	0–1	0–1	0–1	0–1	0–1
Word complexity (%) ($X^2 = 246.9, p < .001$)							
Mean	16.3	33.8	8.5	9.6	18.5	8.1	not relevant
Range	0–1	0–1	0–1	0–1	0–1	0–1	

*Bulgarian frequency was determined on the basis of subjective ratings and is presented in raw figures.

words in Bulgarian to 33.8% of the words in German. This last finding does not correspond to any obvious gradient of inflectional complexity, and indeed examination of the target words themselves suggests that inflected forms (diminutives, plurals) were relatively rare. Rather, the cross-language variation in complexity appears to reflect variations in the use of compounds and/or multi-word target names (by this measure, almost all of the words in Chinese would have been scored as *complex*).

Table 20 provides a more detailed breakdown of length in syllables, showing the number and percentage (out of 520) of the target names for each language that fell into each of eight syllable-length bins. It is clear from this table that monosyllables were the modal response in English (comprising 45.7% of the words that

emerged as the dominant response), contrasting with lows of 0.8% in Italian and 2.3% in Spanish. Disyllables were (as was expected) the modal response in Chinese, where they comprise 70% of all target names. Disyllables were also the modal response in German (49.8%), Spanish (43.8%), Bulgarian (53.7%), and Hungarian (51.2%), albeit at lower levels. In Italian, trisyllables were the modal response (40.0%), although disyllables were not far behind (37.9%). To the extent that word length does play a role in picture naming, we may expect this variable to contribute to cross-language differences in naming behavior. Because we anticipated large cross-language differences in syllable type, we constructed a measure of syllable type frequency for each language (see the Method section). Effects of length, syllable type,

Table 20
Number and Percentage of Target Names Within Each Language
Falling Into Each Category of Syllable Length (Out of 520 Items)

No. of Syllables	English	German	Spanish	Italian	Bulgarian	Hungarian	Chinese
Monosyllables	236	118	12	4	53	90	61
	(45.7%)	(22.7%)	(2.3%)	(0.8%)	(10.2%)	(17.3%)	(11.7%)
Disyllables	201	259	228	197	279	266	364
	(38.7%)	(49.8%)	(43.8%)	(37.9%)	(53.7%)	(51.2%)	(70.0%)
3 syllables	63	104	179	208	140	110	86
	(12.1%)	(20.0%)	(34.4%)	(40.0%)	(26.9%)	(21.2%)	(16.5%)
4 syllables	18	36	75	76	35	41	8
	(3.5%)	(6.9%)	(14.4%)	(14.6%)	(6.7%)	(7.9%)	(1.5%)
5 syllables	2	2	19	24	5	9	1
	(0.4%)	(0.4%)	(3.7%)	(4.6%)	(1.0%)	(1.7%)	(0.2%)
6 syllables	0	1	6	6	6	3	0
	(0.0%)	(0.2%)	(1.2%)	(1.2%)	(1.2%)	(0.6%)	(0.0%)
7 syllables	0	0	1	4	2	0	0
	(0.0%)	(0.0%)	(0.2%)	(0.8%)	(0.4%)	(0.0%)	(0.0%)
8 syllables	0	0	0	1	0	1	0
	(0.0%)	(0.0%)	(0.0%)	(0.2%)	(0.0%)	(0.2%)	(0.0%)

and other word structure variables on naming behavior will be described at the end of Part III.

Correlations within languages. Tables 21–27 (on our Web site at http://www.crl.ucsd.edu/~aszekely/ipnp/7lgpno.html) summarize correlations among the predictor variables (a separate table for each language), including the measure of syllable type frequency designed to capture the cross-language differences in word structure that are evident in Table 20.

As was expected, there were significant correlations among the properties of target words. Length in syllables and length in characters were, of course, highly correlated, but there was some variation: The lowest correlation was in German (which has an especially high ratio of consonant clusters), and the highest was in Italian (an orthographically transparent language with very few consonant-final words). Length in syllables was significantly and negatively correlated with syllable type frequency, reflecting the rarity of words with more than four syllables in all of these languages. However, the magnitude of this correlation varied markedly over languages, from a very low correlation in Chinese (where the overwhelming majority of words are disyllables) to a correlation approaching 1.0 in English (a near-perfect relationship between number of syllables and frequency of each word type). As was anticipated, there were significant confounds between length and complexity, although these confounds varied in magnitude from measure to measure and from language to language. In particular, correlations in every language provide evidence for Zipf's law, the long-noted tendency for languages to reserve shorter words for more frequent concepts. However, these relationships also varied in magnitude over languages. The remaining correlations in Tables 21–27 (e.g., with initial frication) were small and largely nonsignificant, but there were enough of them to warrant inclusion in regression analyses.

Goodness-of-depiction ratings and visual complexity were selected to measure properties of the picture stimuli. The relationship between these two pictorial measures was significant but very small, indicating that more complex pictures tend to be rated a bit more highly. Both picture measures also proved to be largely independent of the properties that characterize their dominant names, although there were a few exceptions. Keeping in mind that goodness ratings were provided by English-speaking raters, we note that *good* pictures tended to be named with words that have atypical syllable structure in Spanish and Bulgarian ($p < .05$), with trends in the same direction for Hungarian and Chinese ($p < .10$). This peculiar finding may reflect (at least in part) the borrowing of English-based words and concepts in some of the languages under study (e.g., words like *igloo* and *Eskimo*). Visual complexity was weakly associated with longer words and with atypical word structure in English, Spanish, and Chinese, suggesting that speakers are coming up with ad hoc constructions to describe some of the more complex pictures. Visual complexity was also weakly but positively associated with word complexity in English and in Spanish. Finally, there was a small but significant positive association between visual complexity and word frequency in Chinese. *The primary lesson to take away from these weak and sporadic effects is that picture and word properties are largely independent*, but there are enough small confounds lurking about to justify inclusion of the picture variables in all regression analyses.

Correlations across languages. The last issue under Part II pertains to cross-language correlations for the major lexical variables in this study, another way to approach the question of universal factors in picture naming. Table 28 presents correlations for the various frequency indices. Even though these indices are all quite different (representing subjective ratings in Bulgarian and estimates from corpora that vary in size and modality for the other six languages), all possible cross-correlations were positive, robust, and significant, ranging from a low of +.37 (between Bulgarian and Chinese and between Bulgarian and Spanish) to a high of +.66 (between German and Hungarian). None of the languages was consistently an outlier. *Hence, it seems that common forces are at work in determining target name frequencies across these seven languages, despite the different ways that frequency was assessed.* Following Johnson et al. (1996), we speculate that at least some of the variance in these seven word frequency measures is contributed by aspects of conceptual accessibility or familiarity that all seven cultures have in common and may not be as direct an index of lexical frequency (lemma and/or word form frequency) as investigators sometimes assume when they use measures of this kind. We will return to this point later.

Table 29 summarizes cross-language correlations for word length in number of syllables (lower triangle) and in number of characters (upper triangle). These findings

Table 28
Correlations Among Word Frequency Measures Across Languages

	EN	GE	SP	IT	BU	HU	CH
English	–						
German	+.65‡	–					
Spanish	+.56‡	+.56‡	–				
Italian	+.54‡	+.55‡	+.54‡	–			
Bulgarian	+.42‡	+.43‡	+.37‡	+.53‡	–		
Hungarian	+.57‡	+.66‡	+.53‡	+.57‡	+.48‡	–	
Chinese	+.53‡	+.56‡	+.43‡	+.47‡	+.37‡	+.46‡	–

‡$p < .01$.

Table 29
Correlations Among Length Measures Across Languages

	EN	GE	SP	IT	BU	HU	CH
English	–	+.45‡	+.43‡	+.40‡	+.29‡	+.38‡	
German	+.44‡	–	+.35‡	+.43‡	+.47‡	+.38‡	
Spanish	+.40‡	+.30‡	–	+.53‡	+.39‡	+.28‡	
Italian	+.42‡	+.38‡	+.49‡	–	+.51‡	+.39‡	
Bulgarian	+.30‡	+.42‡	+.35‡	+.45‡	–	+.30‡	
Hungarian	+.36‡	+.39‡	+.28‡	+.40‡	+.31‡	–	
Chinese	+.37‡	+.34‡	+.25‡	+.32‡	+.20‡	+.39‡	–

Note—Upper triangle, length measured in characters; lower triangle, in syllables. ‡$p < .01$.

are even more surprising than the results for frequency given above: All possible cross-correlations of word length were positive and robust ($p < .01$), ranging from a low of +.20 for length in syllables between Bulgarian and Chinese to a high of +.53 for length in characters between Spanish and Italian. The cross-language correlations for frequency given above were explainable in terms of cross-language commonalities in the familiarity or accessibility of the pictured concepts. But why should we find any correlations over languages in word length? We suggest that this result is a secondary effect of Zipf's law: Short words tend to be reserved for more frequent concepts, and conceptual frequency is shared across the seven languages under study here (we will return to this point later, in regression analyses within and across languages).

Finally, Table 30 presents the cross-correlations among these languages for initial frication (a binary variable). Some of these correlations failed to reach significance, and those that did reach significance were modest in size, as compared with the frequency and length coefficients reported above. But all of the significant correlations were in the same direction, ranging from a low of +.09 (between Spanish and Bulgarian) to a high of +.33 (between English and German). The real surprise here is the fact that there were any significant correlations at all! Why should these seven languages, from a range of different language families, tend to assign target names with the same (or similar) initial consonant? Both qualitative and quantitative assessment of the 520 target names for each language (details at http://www.crl.ucsd.edu/~aszekely/ipnp/7lgpno.html) provides an answer: All of these languages share at least a few cognates, including not only the words that evolved from common ancestors

(e.g., Germanic or Romance), but also some words borrowed directly from English (the base language for which the initial set of picture stimuli was derived). It is likely that we would find such modest positive correlations for any phonological or phonetic coding that is sensitive to cognate status. We chose frication not for this purpose, but because initial fricatives can affect detection of RTs by a voice key. Nevertheless, the correlational result is useful because it alerts us to the many factors that can drive cross-language similarities and differences in naming behavior.

Part II: Interim Summary

To summarize the results for Part II, let us return to the four main questions it posed.

Question II-1. How much will languages differ in the properties of target names that are elicited by the same 520 pictures of common objects?

We found many significant cross-linguistic differences in target words elicited by our common set of picture stimuli. In particular, target names in these seven languages differed significantly across multiple parameters of word structure (length in syllables or orthographic characters, syllable type frequency, and complexity). In view of these differences, it is all the more surprising (and gratifying) that we found so many cross-linguistic universals in naming behavior in the results for Part I.

Question II-2. Will picture properties (objective complexity and goodness-of-depiction) be independent of each other, and will they be independent of the properties of target names within each language? If we do find picture–name confounds (e.g., complex pictures elicit more complex names), will these be language spe-

Table 30
Correlations of Initial Frication (0,1) Across Languages

	EN	GE	SP	IT	BU	HU	CH
English	–						
German	+.33‡	–					
Spanish	+.24‡	n.s.	–				
Italian	+.29‡	+.12‡	+.22‡	–			
Bulgarian	+.25‡	n.s.	+.09†	+.23‡	–		
Hungarian	+.20‡	n.s.	+.08*	+.20‡	+.21‡	–	
Chinese	+.11†	n.s.	n.s.	+.09†	+.08*	n.s.	–

*$p < .1$. †$p < .05$. ‡$p < .01$.

cific, or will they generalize across languages in the study?

Picture properties were largely independent of target word properties in every language, although there were some small confounds that could affect the raw correlations between independent and dependent variables, justifying their inclusion in the multivariate analyses that we are about to present (Part III).

Question II-3. Will we replicate Zipf's law across all the languages in the study? That is, will frequency be associated universally with longer words? And will the same law extend to other properties of word structure (e.g., canonical syllable structure)?

There do appear to be some universal patterns in the relationships among word properties, especially the well-established correlation between length and frequency referred to as Zipf's law. This oft-cited relationship also extended to relationships between frequency and other aspects of word structure that have been examined far less often in the psycholinguistic literature (e.g., word complexity and syllable structure).

Question II-4. Will we find correlations in word characteristics across the seven languages in this study? If the link between meaning and word form is truly arbitrary, we would not expect to find cross-language correlations in the properties of target picture names (e.g., frequency, length, or word structure). To the extent that we do find strong cross-language correlations in word properties (e.g., word frequency in Language A correlates with word frequency in Language B), a case can be made that these correlations reflect universals at a deeper level in the picture-naming process (e.g., conceptual accessibility).

The most surprising finding in Part II lies in the existence and magnitude of the cross-language correlations for all these predictor variables, including weak but significant correlations for initial frication, as well as larger correlations for frequency and length. *These results suggest that lexical frequency and surface characteristics of word structure may reflect processes at a deeper level, rooted in conceptual accessibility, as well as in language history*. This brings us to Part III, where we will examine the effects of lexical and pictorial properties on naming behavior.

Part III: Effects of Word and Picture Properties on Naming Behavior

Part III emphasizes predictor–outcome relationships within and across languages. Because there are significant correlations among many of the predictor variables, we will emphasize regression analyses within each language (e.g., the contributions of frequency, length, initial fricatives, and properties of the picture stimuli to name agreement and naming latencies) and regressions across languages (e.g., the extent to which frequency in Language A predicts naming behavior in Language B). Three critical questions will be addressed in Part III.

Question III-1. What word and picture properties provide the best predictors of naming behavior? Are name agreement and latency affected by the same variables or by different variables? Are the same patterns observed in every language, or do some languages display specific predictor–outcome relationships that are not universally observed?

Question III-2. Will we observe cross-language correlations between predictors and outcomes? For example, will frequency or length in Language A predict RTs in Language B? To the extent that this is the case, we must (again) conclude that these word properties are contaminated by variance at a deeper level in the naming process (e.g., associated variations in picture characteristics and/or conceptual accessibility). We will pursue these particular questions with some novel cross-linguistic variables, including *other-language frequency* versus *own-language frequency*.

Question III-3. What are the contributions of language-specific word structure properties to naming behavior? For example, in languages that favor multisyllabic words, does syllable structure play a role that is not observed in languages in which short words prevail?

Correlations within languages. Tables 31–37 are available at our Web site (at http://www.crl.ucsd.edu/ ~aszekely/ipnp/7lgpno.html), summarizing raw correlations within each language between eight predictor variables and six outcome variables. The eight predictor variables were goodness-of-depiction, visual complexity, word frequency, initial frication, length in syllables, syllable type frequency, length in characters, and word complexity (the latter two are not available for Chinese). Six outcome variables were chosen from the larger set of correlated naming measures because they are partially dissociable and, thus, offer distinct perspectives on naming behavior: percentage of target name agreement (i.e., dominant response, or Lexical Code 1), number of types, same name, target name RT, naming disparity scores (where a positive score indicates relatively higher agreement in that language, as compared with the other six), and target RT disparity scores (where a negative score indicates relatively faster RTs in that language, as compared with the other six).

Because Part II showed substantial collinearity among the predictor variables (within and across languages), these correlational results must be followed up with multiple regressions, summarized below. We refer readers directly to the correlation tables (Tables 31–37) for individual results and will restrict ourselves here to a few summary statements.

First, goodness of depiction and word frequency were the strongest predictors of naming behavior in all languages: With a few exceptions, items with higher picture ratings and higher word frequencies were associated with higher name agreement, fewer alternative names, and faster RTs.

Second, objective visual complexity and initial frication had very few effects on any of the dependent variables, although there were enough small and sporadic effects to justify including these two measures in all regression analyses.

Third, effects of the various length and word structure measures were usually significant but quite modest, and

they varied in magnitude over languages. Because of confounds among the different length variables, these relationships are difficult to interpret (see the comparative analyses of word structure, below). The general tendency is that longer, more complex, and/or less typical word forms (i.e., unusual syllable structure) are associated with lower agreement on target names, more alternative names, more use of shared names, and longer RTs.

Fourth, the evidence in Tables 31–37 also helps to characterize *same names*, a dichotomous variable that captures those cases in which target names are used for more than one item. These shared names tend to be significantly shorter, less complex, more typical in their word structure (i.e., number of syllables), and higher in frequency in every language. This result is compatible with the idea that speakers resort to common, multipurpose word forms when they are "stuck" on difficult items (as reported for English by Székely et al., 2003).

Because of the surprising evidence uncovered in Part II regarding cross-language correlations in frequency and length, we also decided to construct summary variables to reflect universal trends in frequency and length. Two principal-component factor analyses were conducted, one on the frequency measures for all seven languages and another on length in syllables for all seven languages (length in characters and word complexity were not used for this purpose, because these measures are not available for Chinese). In each factor analysis, a single factor emerged with an eigenvalue greater than one, corresponding to *universal frequency* and *universal length*, respectively. Table 38 summarizes correlations of these two factor scores, as well as the shared measures of goodness of depiction and visual complexity, with the five cross-language measures of naming behavior described earlier.

The results were very clear. First, objective visual complexity bore no relationship to any of the cross-language summary variables for naming behavior. Second, even though the ratings were obtained from English raters only, goodness of depiction bore the largest and most consistent relationship to cross-language naming behavior (all *p*s < .01), from a low of −.14 with naming disparity (indicating less disparity over languages for highly rated pictures) to a high of −.57 with average naming latencies (indicating universally faster RTs for highly rated pictures). Third, the cross-linguistic fre-

quency index correlated significantly with all summary variables for naming performance (all *p*s < .01), from a low of −.23 for disparity in name agreement (less disparity for items that elicit high-frequency words across languages) to a high of −.44 for *universal* naming latency (universally faster responses for items that elicit words of greater average frequency). Fourth, the results for *universal length* were weaker than the results for *universal frequency*, but all correlations were significant and in the predicted direction (lower name agreement, slower RTs, and more cross-language disparity for items that elicit words of greater average length).

Regressions of predictors on dependent variables. To disentangle the many confounds among our predictor variables reported in Part II, we turn next to regression analyses of predictor–outcome relationships. These analyses were conducted on the same six dependent variables described above: percentage of target name agreement, number of types, the binary variable *same name*, target RT, naming disparity, and RT disparity. A series of stepwise analyses were conducted on each variable within each language, using the five basic predictor measures that all seven languages share: length in syllables, initial frication, word frequency, goodness of depiction, and visual complexity. Each analysis was repeated five times, permitting each predictor to enter into the equation last, in order to assess its unique contribution to the outcome when the others were controlled. Tables 39–42 summarize the results across languages, with one table for each of the six dependent variables.

Name agreement. As is summarized in Table 39, the five variables together accounted for a significant proportion of the variance in name agreement (*p* < .01) in all seven languages, ranging from a low of 8.0% in Bulgarian to a high of 19.8% in English.

In all the languages, the single largest unique contribution to name agreement came from goodness of depiction, with positive contributions on the last step ranging from a low of +4.9% in Italian to a high of +14.5% in English (the language in which these ratings were derived). The direction of these findings is not surprising: Name agreement tends to be higher for those pictures that are rated as good representations of the English target name. However, the fact that ratings by English speakers work so well across widely different languages did come as a surprise, suggesting that *English goodness-of-*

Table 38
Correlations of Independent With Dependent Cross-Language Summary Variables

	Average Name Agreement	Average Target Latencies	Average Number of Types	Disparity in Name Agreement	Disparity in Target Latencies
Cross-language length factor	−.11†	+.18‡	+.10†	+.19‡	+.20‡
Cross-language frequency factor	+.25‡	−.44‡	−.27‡	−.23‡	−.35‡
Goodness of depiction	+.41‡	−.57‡	−.50‡	−.14‡	−.25‡
Visual complexity	n.s.	n.s.	n.s.	n.s.	n.s.

†*p* < .05.　‡*p* < .01.

Table 39
Regressions of Five Major Independent Variables on Name Agreement

Predictors	English	German	Spanish	Italian	Bulgarian	Hungarian	Chinese
Goodness of depiction	+14.5‡	+6.9‡	+6.1‡	+4.9‡	+7.0‡	+6.6‡	+10.5‡
Visual complexity	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Frequency	+3.2‡	+3.4‡	+0.8†	+2.9‡	+0.8†	n.s.	+6.2‡
Length in syllables	n.s.	n.s.	−1.4‡	−1.8‡	n.s.	−1.5‡	+0.9†
Initial frication	+0.5*	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Total	19.8‡	11.4‡	9.2‡	11.4‡	8.0‡	10.1‡	19.2‡

Note—Total variance and percentage of unique variance accounted for by each predictor after all the others are controlled. Plus/minus indicates direction of contribution. $*p < .1$. $†p < .05$. $‡p < .01$.

depiction ratings may reflect a relationship between the picture and its associated concept, rather than between the picture and its English target name.

In all of the languages except Hungarian, there were also unique positive contributions of word frequency on the last step, indicating higher name agreement for more frequent words. There were no significant contributions of visual complexity or initial frication in any language when these variables were entered on the last step.

The most uneven findings for name agreement were those involving length in syllables: This variable made a unique *positive* contribution to name agreement in Chinese (+0.9%, $p < .05$) but a unique *negative* contribution to name agreement in Spanish (−1.4%, $p < .01$), Italian (−1.8%, $p < .01$), and Hungarian (−1.5%, $p < .01$), whereas it made no significant contribution at all to name agreement in English, German, or Bulgarian. We suspect that these variations are influenced by the large differences in word structure, including word complexity and typicality of syllable structure (as described above). We will return to this issue shortly in more detailed analyses that compare all length-related word structure measures.

Number of types. As is summarized in Table 40, the five measures together also accounted for a significant amount of variance in number of types in every language ($p < .01$) and at levels consistently higher than the corresponding values for name agreement. The total variance ranged from a low of 13.3% in Italian to a high of 26.3% in English.

In every language, the strongest unique contribution to number of alternative names came from goodness of depiction. These ratings made large and robust negative

contributions on the last step, ranging in magnitude from a low of −8.8% in Italian to a high of −22.1% in English (which is, of course, the language in which these ratings were derived). *In other words, fewer alternative names were produced in every language for those pictures that were judged as good representations of English target names.*

The unique contribution of frequency to alternative names failed to reach significance in Bulgarian or Hungarian. In the other five languages, frequency contributions on the final step were significant, ranging from a low of −1.4% ($p < .05$) in Spanish to a high of −6.2% ($p < .01$) in Chinese. Hence, there is an (almost) universal tendency to produce fewer alternative names when the target name is higher in frequency.

There were no unique contributions of visual complexity or initial frication to alternative names in any language. There were also very few significant effects of length in syllables to number of types, restricted to small but significant contributions on the last step in Spanish (+1.1, $p < .05$) and Italian (+0.9, $p < .05$), suggesting that alternative names are used more often in these languages if the target name is relatively long. To some extent, this may reflect our instructions to participants, who were asked to produce the shortest name that came to mind, preferably a single word. However, we suspect that these results are, in part, reflections of the variations in word structure (including typical syllable length), explored in more detail below.

Same names. The results of regression analyses on this variable are summarized in Table 41. They are largely compatible with the notion that *same names* (use of the same word for more than one picture) reflects a

Table 40
Regressions of Five Major Independent Variables on Number of Types

Predictors	English	German	Spanish	Italian	Bulgarian	Hungarian	Chinese
Goodness of depiction	−22.1‡	−13.0‡	−11.1‡	−8.8‡	−13.6‡	−13.5‡	−13.8‡
Visual complexity	n.s.	+0.5*	n.s.	n.s.	n.s.	n.s.	n.s.
Frequency	−1.8‡	−2.2‡	−1.4†	−2.4‡	n.s.	−0.5*	−6.2‡
Length in syllables	n.s.	n.s.	+1.1†	+0.9†	n.s.	n.s.	n.s.
Initial frication	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Total	26.3‡	16.7‡	14.9‡	13.3‡	14.3‡	14.7‡	22.3‡

Note—Total variance and percentage of unique variance accounted for by each predictor after all the others are controlled. Plus/minus indicates direction of contribution. $*p < .1$. $†p < .05$. $‡p < .01$.

Table 41
Regressions of Five Major Independent Variables on Same Name

Predictors	English	German	Spanish	Italian	Bulgarian	Hungarian	Chinese
Goodness of depiction	n.s.	n.s.	n.s.	n.s.	+0.9†	+0.5*	n.s.
Visual complexity	n.s.	−0.7*	n.s.	n.s.	n.s.	+0.5*	−1.5‡
Frequency	+0.8†	+0.5*	+1.7‡	+0.8†	+3.6‡	+3.8‡	−1.1†
Length in syllables	n.s.	−1.1†	−0.6*	n.s.	−2.1‡	n.s.	−3.7‡
Initial frication	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Total	2.8†	4.0‡	3.7‡	1.8 n.s.	7.2‡	7.7‡	7.2‡

Note—Total variance and percentage of unique variance accounted for by each predictor after all the others are controlled. Plus/minus indicates direction of contribution. *$p < .1$. †$p < .05$. ‡$p < .01$.

strategy that speakers use when they find it difficult to come up with a picture description. However, all of these effects are modest. Table 41 shows that the five predictors explained significance variance in same names in six of the seven languages, although these joint contributions were rather small. They ranged from nonsignificance in Italian to a high of 7.7% in Hungarian ($p < .01$).

Recall that the biggest independent contributions to name agreement and number of types came from goodness of depiction. In contrast, goodness of depiction had very little effect on name sharing in any language (see Table 41), suggesting that name sharing occurs not because the picture itself is hard to recognize, but because the concept it represents is difficult to encode. The biggest contributor to name sharing on the final step was word frequency, although these results also varied over languages. Frequency made a small but significant positive contribution on the last step in five languages (English, Spanish, Italian, Bulgarian, and Hungarian). These values ranged from a low of +0.8% ($p < .05$) in English (the language with the lowest incidence of name sharing at 4.5%) to a high of +3.8% ($p < .01$) in Hungarian (which has the second highest incidence of name sharing at 14.0%). These results suggest that speakers tend to resort to certain high-frequency names when they are "stuck" for an answer on difficult items, in line with previous findings for English only (Székely et al., 2003).

However, we were surprised to find a significant contribution in the opposite direction for Chinese (−1.1%, $p < .01$), the language with the highest incidence of name sharing (at 19.6%, close to one fifth of the picture stimuli). That is, shared words in Chinese tend to be lower in frequency when other variables are controlled. This re-

sult suggests that the phenomenon of name sharing may be more complex than we thought in our earlier work on English. The difference may stem from the extensive use of compounding in Chinese, which involves a kind of name sharing at the sublexical level (Lu et al., 2001). That is, a "typical" Chinese noun is made up of two or more sublexical elements, each with an independent meaning (although that meaning is often modified when it is combined with other words). Each of these syllables may occur in many different words, so that their sublexical frequencies are very high. However, for the most common and productive word templates in this language, frequencies at the whole-word level can be relatively low (reflecting a common pattern over languages for highly productive patterns to have high type frequency but low token frequency). When Chinese speakers are trying to find a name for especially difficult pictures, they may use productive compounding to produce an expression that is low in frequency at the whole-word level even though it is made up of common (generic) elements that are easy to retrieve at the sublexical level. This would be similar to a situation in which English speakers produce constructions such as *flying machine* when they are unable to retrieve a target word such as *helicopter*.

Most of the remaining effects on shared names were small and variable, reflecting cross-linguistic variations in word structure that we will consider in more detail later.

Naming latencies. Table 42 presents joint and unique contributions of the same five variables to target-naming latencies. In contrast with the modest to moderate results obtained in comparable analyses of name agreement, number of types, and same names, these five predictor variables explained a substantial amount of the variance in naming time in every language. Values ranged from a

Table 42
Regressions of Five Major Independent Variables on Target Reaction Time

Predictors	English	German	Spanish	Italian	Bulgarian	Hungarian	Chinese
Goodness of depiction	−31.7‡	−22.8‡	−19.8‡	−19.8‡	−24.5‡	−20.6‡	−24.6‡
Visual complexity	n.s.	+0.6†	n.s.	n.s.	+0.6†	n.s.	n.s.
Frequency	−6.8‡	−5.6‡	−3.2‡	−8.2‡	−7.1‡	−3.6‡	−10.1‡
Length in syllables	n.s.	n.s.	+1.0‡	+0.6†	n.s.	+0.9†	n.s.
Initial frication	n.s.	+0.8†	n.s.	n.s.	+0.7†	n.s.	n.s.
Total	43.3‡	33.5‡	26.3‡	30.8‡	32.4‡	28.4‡	39.6‡

Note—Total variance and percentage of unique variance accounted for by each predictor after all the others are controlled. Plus/minus indicates direction of contribution. †$p < .05$. ‡$p < .01$.

low of 26.3% of the variance in Spanish to a high of 43.3% in English.

In every language, the largest contributions by far came from goodness-of-depiction ratings. When these ratings were entered into the equation on the last step, they made negative contributions ranging from −19.8% in Spanish and Italian ($p < .01$) to −31.7% in English ($p < .01$). In other words, *pictures that were judged by English raters to be better representations of their English target names were associated with substantially faster reaction times in every language in the study.*

The unique contributions of word frequency to naming times were smaller but quite robust in every language, ranging from −3.2% in Spanish to −10.1% in Chinese (all $ps < .01$). Hence, we may conclude that the facilitative effect of word frequency on picture-naming latencies is a cross-language universal. This is true despite wide differences in the methods by which word frequency was assessed (although, as we shall see below, the locus of this effect is far from clear).

The remaining effects were small and sporadic, varying from one language to another. Visual complexity made small positive contributions to naming latencies in German and Bulgarian (+0.6%, $p < .05$ in both cases), suggesting that more complex pictures take longer to name. But the same effect did not reach significance in the other languages. Initial frication was also associated with significantly slower RTs in German (+0.8%, $p < .05$) and Bulgarian (+0.9%, $p < .05$), but this effect also failed to reach significance in the other languages. It is not at all obvious to us why these two languages were the only ones to show the anticipated slowing of voice detection for words that begin with white noise. Although German and Bulgarian both have relatively high proportions of fricative-initial words (27.5% and 27.1%, respectively), proportions are even higher in English (28.1%) and Hungarian (34.8%), which did not show this effect. The answer may lie in idiosyncrasies of word structure that have escaped us here but are related indirectly to initial frication. For example, fricatives are used in some languages in the formation of consonant clusters, which are often present in words with complex derivational morphology and (by extension) complex semantics. Finally, there were small but significant unique contributions of length in syllables for only three of the

seven languages: +1.0% ($p < .01$) in Spanish, +0.6% ($p < .05$) in Italian, and +0.9% ($p < .05$) in Hungarian. These were all in the expected direction, with longer RTs for longer words. But this relationship is clearly not universal, at least not in its simplest form. That is, longer words do not always take longer to produce. As we shall see shortly, results for length vary over languages, depending on the measure of length and word structure that we use.

Predictors of cross-language disparity. For the four dependent variables discussed so far, regressions yielded information about universal effects on naming behavior. For the two measures of disparity that we turn to next, regressions will tell us instead about the factors that lead to language differences—that is, to a relative advantage or disadvantage for *each* language, as compared with the other six. The results for regressions on naming disparity are presented in Table 43; the results for disparity in RTs are presented in Table 44. In contrast with the relatively strong and consistent findings across Tables 39–42, few of the results in Tables 43 and 44 reached significance. *In other words, these five predictors are not very effective in explaining language differences, even though they are quite effective in explaining language universals* (see Tables 43 and 44 for details).

Universal predictors. We also conducted a series of *universal* regressions, using four cross-language predictors (goodness of depiction, visual complexity, cross-language frequency, and cross-language length) and five cross-language summary variables for naming behavior (averages over languages for name agreement, number of types, and naming latency; averages of the absolute values for disparity scores in name agreement and naming latency). Details for these analyses are summarized in Table 45. For all five cross-linguistic dependent variables, the four cross-language predictor variables accounted together for a robust and significant proportion of the variance, ranging from lows of 18.2% ($p < .01$) for both of the disparity scores to a high of 49% ($p < .01$) in the cross-language average for naming RTs.

Goodness of depiction emerged once again as the strongest universal predictor of naming behavior (not surprising, since it was also the strongest predictor within each individual language). This variable was associated with universally higher name agreement (+15.9%), fewer

Table 43
Regressions of Five Major Independent Variables on Naming Disparity Scores for Each Language

Predictors	English	German	Spanish	Italian	Bulgarian	Hungarian	Chinese
Goodness of depiction	+2.2‡	n.s.	n.s.	−1.1†	n.s.	n.s.	n.s.
Visual complexity	n.s.	−0.7*	n.s.	n.s.	n.s.	n.s.	n.s.
Frequency	n.s.	+0.8†	n.s.	+1.0†	n.s.	n.s.	+1.6‡
Length in syllables	n.s.	+0.5*	−1.3†	−2.0‡	n.s.	−0.6*	+0.9†
Initial frication	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Total	2.8†	1.9*	1.6 n.s.	5.4‡	0.7 n.s.	1.4 n.s.	3.4‡

Note—Total variance and percentage of unique variance accounted for by each predictor after all the others are controlled; positive score indicates a relative advantage for that language. Plus/minus indicates direction of contribution.   *$p < .1$.   †$p < .05$.   ‡$p < .01$.

Table 44
Regressions of Five Major Independent Variables on
Target Reaction Time Disparity Scores for Each Language

Predictors	English	German	Spanish	Italian	Bulgarian	Hungarian	Chinese
Goodness of depiction	−2.4‡	n.s.	n.s.	+0.9†	n.s.	n.s.	n.s.
Visual complexity	−0.6*	+0.7*	n.s.	n.s.	+0.8†	n.s.	n.s.
Frequency	n.s.	n.s.	n.s.	−2.1‡	n.s.	n.s.	−1.4‡
Length in syllables	n.s.	n.s.	+0.7*	+1.6‡	+0.8†	+0.6*	n.s.
Initial frication	n.s.	+1.5‡	n.s.	n.s.	+2.4‡	n.s.	n.s.
Total	3.5‡	2.3†	1.6 n.s.	6.4‡	3.9‡	1.0 n.s.	2.1*

Note—Total variance and percentage of unique variance accounted for by each predictor after all the others are controlled; negative score indicates a relative advantage for that language. Plus/minus indicates direction of contribution. *$p < .1$. †$p < .05$. ‡$p < .01$.

alternative names (−23.6%), and faster RTs for the target name (−28.8%). Its effects on the two disparity measures were much smaller, but the direction of these contributions indicates that higher picture ratings are associated with less disparity in name agreement (−1.7%, $p < .01$) and less disparity in RTs (−5.6%, $p < .01$). Hence, the slight advantage that goodness-of-depiction ratings provide for English participants is more than offset by the overall advantage that *good* pictures confer across all seven languages.

Objective visual complexity had no detectable effect on any of these cross-language summary variables, even though it did have sporadic effects within individual languages. Hence, the robust universal effect of goodness of depiction derives not from the objective complexity of these pictures, but from how well each picture represents its associated concept.

Universal frequency (the average of *z* score frequencies across all seven languages) made significant unique contributions on the last step for all five cross-language summary variables, resulting in higher name agreement, fewer alternative names, faster RTs, and less disparity over languages in both agreement and RT.

Finally, the cross-language summary score for target word length made a modest but significant contribution to disparity in name agreement (+1.1%, $p < .01$), indicating that there is more disparity in naming for those items that tend to elicit longer descriptions. Universal length made no other unique contributions on the last step, which suggests to us that cross-language correlations in length are indeed (as we suggested above) re-

flections of Zipf's law: When the confound between *universal length* and *universal frequency* is controlled in this regression analysis, the *universal* effects of length are (with the single exception just noted) largely eliminated. Note that this conclusion pertains only to universal cross-language effects. As we will see shortly, language-specific effects of length and other aspects of word structure do sometimes persist within specific languages, surviving even after frequency is controlled.

To summarize the results so far for Part III, correlation and regression analyses involving picture properties indicate that goodness of depiction is the single best predictor of naming behaviors in every language, even after other variables are controlled. This is true despite the fact that goodness ratings (which were collected in English) are associated with a slight advantage for English speakers (as reflected in naming and RT disparity scores). In contrast, objective visual complexity has very few effects on naming behavior and no effect at all at the level of cross-language summary variables. It is associated with some small and sporadic effects within individual languages, but the direction of those effects can vary (i.e., greater detail is sometimes associated with more name agreement, but greater complexity is occasionally associated with slower RTs). In general, we may conclude that *picture difficulty* is a cognitive effect that is captured reasonably well by goodness-of-depiction ratings and appears to be largely independent of visual complexity, at least as we have measured it here.

With regard to the predictive value of word properties, word frequency is the second-best predictor of naming

Table 45
Regressions of Cross-Language Independent Variables on Cross-Language Summary Scores for
Naming Behavior: Total Percentage of Variance Accounted for and Unique Contributions of Each
Predictor on the Final Step

Predictors	Average Name Agreement	Average Target Latencies	Average Number of Types	Disparity in Name Agreement	Disparity in Target Latencies
Goodness of depiction	+15.9‡	−28.8‡	−23.6‡	−1.7‡	−5.6‡
Visual complexity	n.s.	n.s.	n.s.	n.s.	n.s.
Cross-language length factor	n.s.	n.s.	n.s.	+1.1‡	n.s.
Cross-language frequency factor	+3.6‡	−12.3‡	−4.3‡	−1.9‡	−7.1‡
Total variance accounted for	22.6‡	49.0‡	30.3‡	18.2‡	18.2‡

‡$p < .01$.

behavior (after goodness of depiction), resulting in higher agreement, fewer alternative names, and faster RTs within and across languages. What remains to be seen is whether these frequency effects are language specific (reflecting the "true" frequency of target names within each individual language) or whether they reflect aspects of conceptual familiarity or accessibility that hold up across languages and are relatively independent of the target words themselves. To pursue this issue further, we turn now to some specific comparisons of *own-language frequency* and *other-language frequency*. For these analyses, we will restrict our attention to the two most important dependent variables in the study: name agreement and latencies to produce the target name.

Cross-language word frequency effects. In Part II, we uncovered some surprisingly high cross-language correlations among our word frequency measures, despite marked differences in the way that frequency norms were collected in each language. This result leads us to ask whether word frequency effects on naming behavior are language specific or whether they are, in fact, interchangeable measures of a common source of variance, including universal variations in the familiarity and accessibility of the concepts illustrated in our 520 pictures.

To investigate this issue, we constructed a novel set of *other-frequency* variables. For each language, a factor analysis was conducted on the frequency measures for the other six languages, and the first principal component from that analysis was used as a single estimate of *other-language frequency*. For example, the other-language frequency factor for English is based on the first principal component from factor analyses of frequency measures for German, Spanish, Italian, Bulgarian, Hungarian, and Chinese. In each of these factor analyses (one for each language), only one factor emerged with an eigenvalue greater than one, which testifies further to the hypothesis that a single underlying construct (e.g., conceptual accessibility) is driving these effects. Table 46 summarizes the cross-language correlations of word frequency with name agreement within and across all seven languages. Table 47 provides corresponding statistics for the cross-language relationship of word frequency to naming times. In both tables, we include correlations with *universal frequency* (the first principal component for all seven frequency measures) as well as *other-language frequency* (the first principal component for the six remaining languages when the target language is excluded).

For name agreement, 54 of the 63 correlations in Table 46 were significant, and all of them were positive, representing higher name agreement for items with more frequent names. Most important for our purposes here, within-language correlations (frequency in Language A with name agreement in Language A) were often lower than cross-language correlations (frequency in Language A with name agreement in Language B). In English, German, and Chinese, the own-language correlation was numerically *larger* than both the other-language correlation and the universal frequency correlation. In the remaining four languages (Spanish, Italian, Bulgarian, and Hungarian), own-language frequency actually resulted in numerically *smaller* correlations than both universal frequency and other-language frequency.

Corresponding results for naming times were even stronger (Table 47): All 63 within- and across-language correlations were significant (at $p < .01$), all were larger than the corresponding values for name agreement, and all were in the negative direction (indicating faster latencies for higher frequency items). In every language except Chinese, the other-language frequency and universal frequency correlations were higher than own-language frequency. To some extent, these results may reflect weaknesses in the frequency measures for some of the individual languages. For example, Bulgarian frequency estimates are based on subjective ratings, and the Spanish, Italian, Bulgarian, and Hungarian frequency corpora are relatively small, as compared with the corpora on which English, German, and Chinese frequencies are based. Nevertheless, *the magnitude and consistency of all the cross-language frequency effects suggests to us that picture-naming behavior is driven, at least in part, by facts about conceptual accessibility and familiarity that are shared across languages, independent of the target name that is actually used.*

A more rigorous test of this hypothesis can be obtained with regression analyses in which the unique contributions of own-language frequency and other-language frequency are compared. Tables 48 and 49 present results of these regressions for name agreement and target RTs, re-

Table 46
Correlations of Word Frequencies With Name Agreement Within and Across Languages

Frequencies From	English	German	Spanish	Italian	Bulgarian	Hungarian	Chinese
English	+.21‡	+.17‡	+.22†	+.29‡	+.16‡	+.21‡	+.09†
German	+.08†	+.20‡	+.18‡	+.25‡	+.13‡	+.17‡	+.07*
Spanish	+.12‡	+.10‡	+.14‡	+.19‡	n.s.	+.16‡	+.11‡
Italian	+.08†	+.11‡	+.15‡	+.22‡	+.09†	+.10†	+.06*
Bulgarian	n.s.	+.07*	+.16‡	+.10†	+.09†	n.s.	+.06*
Hungarian	+.08†	+.11‡	+.13‡	+.19‡	+.06*	+.12‡	n.s.
Chinese	+.13‡	+.13‡	+.20‡	+.21‡	+.10†	+.20‡	+.26‡
Other-language frequency	+.11‡	+.15‡	+.22‡	+.27‡	+.12‡	+.19‡	+.09†
Universal frequency	+.14‡	+.17‡	+.22‡	+.28‡	+.13‡	+.19‡	+.12‡

*$p < .1$.   †$p < .05$.   ‡$p < .01$.

Table 47
**Correlations of Word Frequencies From Each Language
With Naming Latencies Within and Across Languages**

Frequencies From	English	German	Spanish	Italian	Bulgarian	Hungarian	Chinese
English	−.34‡	−.31‡	−.37‡	−.39‡	−.28‡	−.29‡	−.33†
German	−.30‡	−.32‡	−.35‡	−.38‡	−.27‡	−.32‡	−.29‡
Spanish	−.27‡	−.28‡	−.24‡	−.33‡	−.21‡	−.26‡	−.25‡
Italian	−.28‡	−.22‡	−.27‡	−.33‡	−.18‡	−.20‡	−.19‡
Bulgarian	−.31‡	−.25‡	−.34‡	−.32‡	−.27‡	−.22‡	−.28‡
Hungarian	−.27‡	−.27‡	−.27‡	−.33‡	−.22‡	−.27‡	−.25‡
Chinese	−.31‡	−.29‡	−.35‡	−.34‡	−.26‡	−.28‡	−.39‡
Other-language frequency	−.38‡	−.35‡	−.42‡	−.45‡	−.30‡	−.33‡	−.34‡
Universal frequency	−.39‡	−.36‡	−.41‡	−.45‡	−.31‡	−.35‡	−.37‡

†$p < .05$.   ‡$p < .01$.

spectively. In these analyses, we also controlled for goodness of depiction, visual complexity, initial frication, and length in syllables, to maximize comparability to the regressions reported earlier in Part III. Each analysis was repeated six times, allowing each predictor an opportunity to enter into the equation last. Although there were a few small changes for other variables (e.g., some small effects that disappeared with the addition of a new variable), for present purposes we will restrict our attention to the unique contributions of own- versus other-language frequency (the reader may refer to Tables 48 and 49 for details regarding the other measures).

For name agreement, own-language frequency made significant and unique positive contributions on the last step in English (+3.3%), German (+1.6%), and Chinese (+6.5%; all ps < .01). For the remaining languages, the effects of own-language frequency were wiped out when other-language frequency was entered into the equation first. Note that English, German, and Chinese are the languages in which frequencies are based on especially large written corpora, which may provide more reliable estimates of word frequencies in their respective languages.

In contrast, contributions to name agreement on the last step for other-language frequency varied over languages in direction, as well as in magnitude. Significant *positive* contributions of other-language frequency were observed on the last step in Spanish (+2.4%), Italian (+1.9%), and Hungarian (+1.6%; all ps < .01). In these three languages, these were the only frequency effects ob-

served. In Bulgarian, no significant effects of *either* own- or other-language frequency were observed, suggesting that these two indices were explaining the same variance in this language. It is interesting to note in this regard that Bulgarian is the only language in which frequency was assessed by subjective ratings. The most peculiar results were observed in Chinese: When own-language frequency was entered into the equation first, a small but significant negative relationship emerged with other-language frequency (−0.7, $p < .05$). In other words, for Chinese speakers, name agreement was actually somewhat *lower* for pictures that were named with high-frequency words in other languages. This may reflect a variety of factors, including a competition between production of Chinese terms and production of foreign loan words that are frequent in some of the other European languages but somewhat less accessible in Chinese.

In the corresponding analyses of naming times, results were stronger and more consistently in favor of other-language frequencies. When own-language frequency was entered into the equation last, significant negative contributions emerged for English (−0.7%), Chinese (−4.1%), and Bulgarian (−2.1%). The emergence of Bulgarian in the RT analyses is interesting, since this is the only language in which frequency was measured by subjective ratings. It suggests that frequency ratings (as opposed to frequency counts) may have a degree of "staying power" for prediction of RTs that they do not have for name agreement in this language.

Table 48
**Regressions on Name Agreement Within Each Language Using Both Own-Language Frequency
and Other-Language Frequency as Predictors**

Predictors	English	German	Spanish	Italian	Bulgarian	Hungarian	Chinese
Goodness of depiction	+14.7‡	+6.9‡	+6.0‡	+4.4‡	+6.7‡	+6.4‡	+10.6‡
Visual complexity	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Length in syllables	n.s.	n.s.	−0.7†	−1.3‡	n.s.	−1.3‡	+0.7†
Initial frication	+.05*	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Own-language frequency	+3.3‡	+1.6‡	n.s.	n.s.	n.s.	n.s.	+6.5‡
Other-language frequency	−.05*	n.s.	+2.4‡	+1.9‡	n.s.	+1.6‡	−0.7†
Total variance	20.3‡	11.5‡	11.5‡	13.3‡	8.4‡	11.7‡	19.9‡

Note—Total variance and percentage of unique variance accounted for by each predictor after all the others are controlled. Plus/minus indicates direction of contribution.   *$p < .1$.   †$p < .05$.   ‡$p < .01$.

Table 49
Regressions on Naming Latencies Within Each Language Using Both Own-Language Frequency and Other-Language Frequency as Predictors

Predictors	English	German	Spanish	Italian	Bulgarian	Hungarian	Chinese
Goodness of depiction	−30.7‡	−23.1‡	−19.2‡	−17.8‡	−23.3‡	−20.0‡	−24.3‡
Visual complexity	n.s.	+0.6†	n.s.	n.s.	+0.7†	n.s.	n.s.
Length in syllables	n.s.	n.s.	n.s.	n.s.	n.s.	+0.6†	n.s.
Initial frication	n.s.	+0.9‡	n.s.	n.s.	+0.6†	n.s.	n.s.
Own-language frequency	−0.7†	n.s.	n.s.	n.s.	−2.1‡	n.s.	−4.1‡
Other-language frequency	−2.8‡	−3.4‡	−10.6‡	−7.3‡	−2.0‡	−3.8‡	−1.4‡
Total variance	46.1‡	36.8‡	36.8‡	38.2‡	34.4‡	32.2‡	41.0‡

Note—Total variance and percentage of unique variance accounted for by each predictor after all the others are controlled. Plus/minus indicates direction of contribution.   †$p < .05$.   ‡$p < .01$.

Most important for our purposes here, *when other-language frequencies were entered into the RT equation last, including a control for own-language frequency, significant unique contributions of other-language frequency were present in all seven languages.* These values were all in the same direction, ranging from a low of −1.4% in Chinese to a high of −10.6% in Spanish (all ps < .01). It is clear from these results that word frequency reflects universal factors that go beyond the frequency of specific word forms within each language.

If it is the case that these cross-language frequency effects apply at a conceptual level, we can make a strong prediction: Facilitating effects of cross-language frequency (and perhaps, own-language frequency) should be observed not only for RTs for producing the target name, but also for the time required to produce *any* legitimate name for the same concept. The number of items receiving Lexical Codes 2 and 3 was relatively small (see Table 4), and some items (in some languages) elicited no instances of morphophonological variants and/or synonyms. Hence, our test of this prediction rests on relatively noisy data, with a smaller number of items. With that caveat in mind, we examined the correlations of own- versus other-language frequencies with RTs for all cases in which speakers produced a response receiving Lexical Codes 2 or 3. These results are presented in Table 50 for each language.

Although many of these correlations failed to reach significance, the pattern in Table 50 is very clear: *RTs to produce an alternative name are faster if the target name is high in frequency.* This result is most evident for other-language frequencies, where it holds for five out of

seven languages in Lexical Code 2 (morphophonological variants) and six out of seven languages on Lexical Code 3 (synonyms with no morphological overlap). But the results are in the same direction for own-language frequencies, where they reach significance for three out of seven languages on Lexical Code 2 and two of the seven languages on Lexical Code 3. In fact, of the 28 possible correlations in Table 50, 17 are significant, and 24 are in the predicted negative direction. This is the opposite of what we would expect if frequency effects in picture naming occur at the level of name selection. That is, *a high-frequency target name ought to inhibit its competitors, and yet we find that any name is easier to produce for those pictures that elicit a high-frequency dominant response.* These results suggest that at least some of the variance in the frequency/RT relationship is due to the familiarity and accessibility of the pictured concept, independently of the name that speakers finally retrieve and produce for that concept.

It has been suggested to us that at least some of the other-language frequency effects reported here might be an indirect reflection of cognates—that is, words that share similarity of form, as well as meaning, across two or more languages. Specifically, it has been suggested that cross-language frequency effects may be driven by words that are *the same* on a variety of levels. The definition of a cognate is neither straightforward nor uncontroversial. Some cases are obvious (e.g., *television* in English and *televisione* in Italian), but others are more opaque and can be recognized only by speakers who know the phonological and morphological relationships between the two languages (e.g., *refrigerator* in English

Table 50
Correlations of Own-Language Frequency and Other-Language Frequency With Reaction Times for Morphophonological Variants (Lexical Code 2) and Synonyms (Lexical Code 3)

	English	German	Spanish	Italian	Bulgarian	Hungarian	Chinese
Own frequency							
Lexical Code 2	n.s.	n.s.	n.s.	−.17‡	−.21‡	n.s.	−.22‡
Lexical Code 3	−.16*	−.23‡	n.s.	−.19‡	n.s.	n.s.	−.16*
Other-language frequency							
Lexical Code 2	−.17†	n.s.	−.26‡	−.23‡	−.25‡	−.08*	−.32‡
Lexical Code 3	−.28‡	−.27‡	−.25‡	−.33†	−.28‡	−.21‡	n.s.

*$p < .1$.   †$p < .05$.   ‡$p < .01$.

Table 51
Cognate Status: Physical Similarity Between Target Names in Average Number of Overlapping
Orthographic Trigrams, Across All Pairs of Languages

	EN	GE	SP	IT	BU	HU	CH
English	–						
German	0.65	–					
Spanish	0.48	0.29	–				
Italian	0.56	0.44	1.01	–			
Bulgarian	0.30	0.48	0.31	0.44	–		
Hungarian	0.19	0.30	0.18	0.27	0.33	–	
Chinese	0.02	0.02	0.02	0.01	0.02	0.00	–
Mean overlap with all other languages	0.37	0.36	0.38	0.46	0.31	0.21	0.02

Note—All languages: 30%.

vs. *frigorifero* in Italian). Hence, a comprehensive coding of cognate status would require expert linguistic judgments applied to phonemically coded words, as opposed to automatic coding procedures applied to orthographic representations. Nevertheless, to obtain a rough estimate of between-language overlap among the target names elicited in this study, we quantified cognates by applying a trigram overlap search algorithm (write to us directly at aszekely@crl.ucsd.edu or see our Web site at http://www.crl.ucsd.edu/~aszekely/ipnp/7lgpno.html for details). For any given pair of languages, a score greater than 1 would reflect an average of at least one trigram overlap for each pair of target names (out of 520 pairs), a score of .25 would represent a situation in which one shared trigram overlap was identified for approximately every four pairs of target words, and so forth.

Table 51 summarizes the amount of trigram overlap identified across each pairing of the seven languages in our study, as well as the average overlap each language shared with the other six combined. Keeping in mind the limits of a cognate metric that is based solely on orthography (rather than phonology) and is applied from left to right without regard to morphological structure (e.g., roots vs. inflections) or language history, Table 51 does provide interesting results. Specifically, it provides our first evidence for the metric of language distance that we expected to find (but did not find) with our other naming variables. The highest overlaps were observed between closely related languages (Italian vs. Spanish and English vs. German), and the lowest overlaps involved Hungarian and Chinese (which come from distinct language families). To determine whether this measure of cognate status might affect the other-language frequency effects reported above, we repeated the regressions on RT in Table 49 but added cognate status (amount of interlanguage overlap) to the list of variables that were controlled. All of the significant results for own-language frequency and other-language frequency remained exactly the same. Hence, the cross-language frequency effects reported above are not simply by-products of cognate status. We reiterate, however, that this is an issue that requires a more detailed inquiry. Even for the crude cognate metric employed here, interesting differences are observed across languages in the relationship between cognate status, frequency, and word structure variables (details are provided on our Web site at http://www.crl.ucsd.edu/~aszekely/ipnp/7lgpno.html).

Cross-language and language-specific effects of word structure. In contrast with word frequency, word length appears to play a relatively minor role in naming behavior. However, length is also the only factor that contributes consistently to cross-language disparity. Because length in syllables is only one aspect of language-specific word structure, we will end Part III with some exploratory analyses of *universal length*, followed by more detailed and language-specific comparisons of different measures of word structure.

To investigate the issue of universal length, we followed the model described above for word frequencies and constructed for each language a measure of *other-language*

Table 52
Correlations of Length in Syllables From Each Language
With Name Agreement Within and Across Languages

Length in Syllables	English	German	Spanish	Italian	Bulgarian	Hungarian	Chinese
English	−.09†	n.s.	−.07*	−.15‡	−.07*	−.08†	n.s.
German	−.06*	n.s.	−.08†	−.11‡	−.08†	n.s.	n.s.
Spanish	−.07*	n.s.	−.13‡	−.11‡	n.s.	−.08†	n.s.
Italian	−.06*	n.s.	−.10‡	−.18‡	n.s.	n.s.	n.s.
Bulgarian	n.s.	+.10†	n.s.	−.08†	n.s.	−.15‡	+.08†
Hungarian	−.10†	n.s.	−.08†	−.15‡	−.09†	−.15‡	n.s.
Chinese	n.a.	n.s.	n.s.	−.15‡	n.s.	−.08†	n.s.
Other-language length	−.08†	n.s.	−.09†	−.19‡	n.s.	n.s.	n.s.
Universal length	−.09†	n.s.	−.11‡	−.20‡	−.08*	−.09†	n.s.

*$p < .1$.   †$p < .05$.   ‡$p < .01$.

Table 53
Correlations of Length in Syllables From Each Language on Naming Latencies Within and Across Languages

Length in Syllables	English	German	Spanish	Italian	Bulgarian	Hungarian	Chinese
English	+.16‡	+.18‡	+.12‡	+.19‡	+.15‡	+.13‡	+.12‡
German	+.18‡	+.16‡	+.14‡	+.13‡	+.14‡	+.11‡	+.14‡
Spanish	+.12‡	+.14‡	+.13‡	+.16‡	+.08†	+.09†	+.06*
Italian	+.19‡	+.13‡	+.16‡	+.14‡	n.s.	n.s.	n.s.
Bulgarian	+.15‡	+.14‡	+.08†	n.s.	n.s.	n.s.	n.s.
Hungarian	+.13‡	+.11‡	+.09†	n.s.	n.s.	+.18‡	+.12‡
Chinese	+.14‡	+.19‡	+.08†	+.19‡	+.13‡	+.12‡	+.17‡
Other-language length	+.14‡	+.14‡	+.14‡	+.22‡	+.16‡	+.11†	+.11‡
Universal length	+.16‡	+.15‡	+.15‡	+.22‡	+.15‡	+.13‡	+.13‡

*$p < .1$.   †$p < .05$.   ‡$p < .01$.

length. That is, we conducted principal-component factor analyses on length in syllables in the six remaining languages when the target language was removed. So, for example, the other-language length measure for Italian was the first principal component for factor analyses of length in syllables in English, German, Spanish, Bulgarian, Hungarian, and Chinese. In all seven of these analyses, only one factor emerged with an eigenvalue greater than 1, providing further evidence that some kind of unitary latent variable underlies all these measures of length. Table 52 summarizes correlations with naming agreement in each language, for the individual length measures in every language, for the universal length factor described in Part II, and for all seven other-language length factor scores. Table 53 presents the corresponding correlations for naming latency. The results were much weaker than the analogous findings for word frequency (in Tables 46 and 47), both within and across languages, and the findings were especially weak for name agreement. Nevertheless, the general trend in Tables 52 and 53 confirms that effects of length on naming behavior are often just as strong across languages (e.g., length in Language A with RTs in Language B) as they are within languages (e.g., length in Language A with RTs in Language A). Hence, there does appear to be a universal factor underlying the relationship between length and naming behavior.

In the explorations of frequency described above, we carried out regression analyses comparing the unique contributions of *own-language frequency* versus *other-language frequencies*. We decided not to pursue such a strategy for *own-language length* versus *other-language length*, because we already know from analyses reported earlier (Table 45) that most of the effects of universal length are wiped out when universal frequency is controlled. In other words, the cross-linguistic component in our length measures seems to be a by-product of the well-known association between length and frequency (Zipf's law). However, there was one important exception to this generalization (in Table 45): Cross-language length did make a significant unique contribution to disparity in name agreement. Hence, variations in word length can lead to language-specific advantages or disadvantages in name retrieval.

Within-language effects of word structure. The remainder of Part III is devoted to a more detailed exploration of language-specific effects involving four partially dissociable measures of word structure: length in syllables, syllable type frequency, length in characters, and word complexity (the latter two measures are not available for Chinese). Although these effects are small, they comprise some of the most consistent findings in this study for cross-linguistic differences in lexical retrieval. Furthermore, such effects may be magnified in developmental and/or clinical studies of word production under nonoptimal conditions. Hence, we believe that they merit consideration.

First, we conducted two rounds of regressions, one for name agreement (Table 54) and the other for target naming times (Table 55), examining the contribution of each of the four word structure variables after goodness of depiction, visual complexity, initial frication, and frequency are controlled. These analyses can tell us which of the four predictors (if any) does the best job of capturing length-related variance for each language when it is working as the only length-related variable. Tables 54 and 55 show

Table 54
Unique Variance Contributed by Each Length Predictor to Name Agreement on the Last Step

Predictors	English	German	Spanish	Italian	Bulgarian	Hungarian	Chinese
Length in syllables	n.s.	n.s.	−1.4‡	−1.8‡	n.s.	−1.5‡	+0.9†
Syllable frequency type	n.s.	n.s.	+1.5‡	+1.0†	n.s.	+2.4‡	+0.6*
Length in characters	−0.8†	n.s.	−1.3‡	−1.8‡	−0.7†	−1.5‡	n.a.
Word complexity	−1.5‡	n.s.	−3.6‡	−2.4‡	n.s.	−2.7‡	n.a.

Note—Each in separate regressions, after goodness of depiction, visual complexity, word frequency, and initial frication are controlled. Plus/minus indicates direction of contribution.   *$p < .1$.   †$p < .05$.   ‡$p < .01$.

Table 55
Unique Variance Contributed by Each Length Predictor to Naming Latencies on the Last Step

Predictors	English	German	Spanish	Italian	Bulgarian	Hungarian	Chinese
Length in syllables	n.s.	n.s.	+1.0‡	+0.6†	n.s.	+0.9†	n.s.
Syllable frequency type	n.s.	n.s.	−1.9‡	−1.2‡	−0.5†	−1.4‡	−2.0‡
Length in characters	+0.5†	n.s.	+1.0‡	+0.6†	+1.1‡	+1.6‡	n.a.
Word complexity	n.s.	n.s.	+2.0‡	n.s.	n.s.	+0.9†	n.a.

Note—Each in separate regressions, after goodness of depiction, visual complexity, word frequency, and initial frication are controlled. Plus/minus indicates direction of contribution.   †$p < .05$.   ‡$p < .01$.

that all of these length-related findings are relatively modest in size. Their primary interest lies in the considerable variation observed over languages in the presence, direction, and magnitude of word structure effects. For example, naming behavior in German (both agreement and latency) seems to be impervious to any of these word structure variables after other factors are controlled. For English and Bulgarian, a few word structure effects reach significance, but they are weak and sporadic. In contrast, almost all of the word structure variables made a contribution on the last step for Spanish, Italian, and Hungarian.

Since word structure appears to be important for at least some of these languages, we pursued the issue further by conducting regressions in which both length in syllables and frequencies of syllable type were used as length-related predictors after other variables were controlled (i.e., goodness of depiction, visual complexity, frequency, and initial frication). To the best of our knowledge, this is the first time that effects of word structure typicality have been assessed in a cross-linguistic framework in any modality including picture naming. Each of these word structure measures (which are available for all seven languages) was given an opportunity to enter into the equation last. The results (summarized in Tables 56 and 57 for name agreement and RTs, respectively) provide further evidence for language specificity.

For name agreement, these length-related measures either canceled each other out or had no effect to begin with in English, German, and Bulgarian. In contrast, Italian showed a small negative effect of length (-0.8%, $p < .05$), whereas Hungarian showed a small positive effect of syllable type ($+1.3\%$, $p < .01$). Chinese showed yet another pattern, with facilitative effects on name

agreement from length ($+1.2\%$, $p < .01$), as well as from syllable type ($+0.9\%$, $p < .05$). As we have noted earlier, the helpful effects of length in Chinese probably reflect the overwhelming predominance of disyllables in that language.

For naming latency, none of the languages showed inhibitory effects of length after all the other variables (including syllable type) were controlled. However, facilitative effects of syllable type were observed in Spanish (-0.8%, $p < .05$), Italian (-0.6%, $p < .05$), Hungarian (-0.7%, $p < .05$), and especially Chinese (-2.0%, $p < .01$). Although these effects are small, they indicate that it is easier in some languages to retrieve words that correspond to the dominant word structure template. This is a novel finding that could prove important in future cross-language studies.

To push the limits of our findings on word structure, one final round of analyses was conducted in which all four word structure variables were used together in the same equation for all languages except Chinese (for which length in characters and word complexity measures were not available). These analyses are summarized in Tables 58 and 59. Once again, each word structure measure was given an opportunity to enter into the equation last, after controlling for the other word structure measures and for goodness of depiction, visual complexity, initial frication, and frequency. Because it would be difficult for any variable to make a unique contribution after this large list of factors is entered into the equation, we also ran a set of analyses in which the four word structure variables were entered together as a block on the last step, to see if their joint contribution is greater in some languages than it is in others. These results are also included in Tables 58 and 59. *The main message to take*

Table 56
Regressions on Name Agreement Using Both Length in Syllables and Syllable Type Frequency

Predictors	English	German	Spanish	Italian	Bulgarian	Hungarian	Chinese
Goodness of depiction	+14.5‡	+7.0‡	+6.3‡	+4.9‡	+7.0‡	+7.0‡	+10.8‡
Visual complexity	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Initial frication	+0.5*	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Word frequency	+3.2‡	+3.5‡	+0.8†	+2.9‡	+0.8†	+0.5*	+6.3‡
Length in syllables	n.s.	n.s.	n.s.	−0.8†	n.s.	n.s.	+1.2‡
Syllable type frequency	n.s.	n.s.	n.s.	n.s.	n.s.	+1.3‡	+0.9†
Total	19.8‡	11.6‡	9.4‡	11.4‡	8.1‡	11.4‡	20.1‡

Note—Total variance and percentage of unique variance accounted for by each predictor on last step. Plus/minus indicates direction of contribution.   *$p < .1$.   †$p < .05$.   ‡$p < .01$.

Table 57
Regressions on Target Reaction Time Using Both Length in Syllables and Syllable Type Frequency

Predictors	English	German	Spanish	Italian	Bulgarian	Hungarian	Chinese
Goodness of depiction	−31.6‡	−22.7‡	−20.4‡	−19.8‡	−24.8‡	−21.0‡	−25.4‡
Visual complexity	n.s.	+0.6†	n.s.	n.s.	+0.6†	n.s.	n.s.
Initial frication	n.s.	+0.8†	n.s.	n.s.	+0.6†	n.s.	n.s.
Word frequency	−6.6‡	−5.2‡	−3.2‡	−8.2‡	−7.1‡	−3.9‡	−10.8‡
Length in syllables	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Syllable type frequency	n.s.	n.s.	−0.8†	−0.6†	n.s.	−0.7†	−2.0‡
Total	43.3‡	33.5‡	27.1‡	31.4‡	32.7‡	29.1‡	41.5‡

Note—Total variance and percentage of unique variance accounted for by each predictor on last step. Plus/minus indicates direction of contribution.   †$p < .05$.   ‡$p < .01$.

away from these analyses is that every language shows a unique pattern of word structure contributions to picture naming.

For name agreement, word structure variables made a significant joint contribution on the last step in English, Spanish, Italian, and Hungarian. For naming latencies, the same word structure variables made significant joint contributions in Spanish, Italian, Bulgarian, and Hungarian. The pattern of specific predictors was different in every language. For example, German proved to be impervious to length-related effects, separately or together, for either dependent variable. For the remaining languages, the best predictor of name agreement on the last step was length in characters for English and Bulgarian and word complexity for Spanish and Italian, whereas syllable type and word complexity had equal effects for Hungarian. For naming latencies, the best predictors on the last step were length in characters for English and Bulgarian, word complexity for Spanish, and syllable type frequency for Italian and Hungarian.

These detailed word structure effects suggest that contributions of word structure to naming behavior are language specific. There are good reasons to believe that such patterns will be even more pronounced in studies of word reading, word recognition, and comprehension—an issue worth pursuing further in studies specifically designed for each language.

Part III: Interim Summary

To summarize the findings for Part III, we return to the three main questions posed for this section earlier on.

Question III-1. What word and picture properties provide the best predictors of naming behavior? Are name agreement and latency affected by the same variables or by different variables? Are the same patterns observed in every language, or do some languages display predictor–outcome relationships that are not universally observed?

In all seven languages, the two best predictors of naming behavior were goodness of depiction (assumed to be a property of the pictures) and word frequency (typically assumed to be a property of the target names). These factors were important for both name agreement and naming latency, but effects were especially strong for latency. In contrast with these *big predictors*, effects of other independent variables were relatively small, tended to vary from one language to another, and were not consistent in size, direction, or the dependent variable on which they had their primary effect. The weak contribution of initial frication is particularly interesting in this regard, because phoneme onset properties are known to have a massive effect on RTs for word naming, which also involves a vocal response (Spieler & Balota, 1997).

Question III-2. Will we observe cross-language correlations between predictors and outcomes? For example, will frequency or length in Language A predict RTs in Language B? If this is the case, what will happen when we compare other-language frequency versus own-language frequency, and other-language length versus own-language length?

We did indeed find robust cross-language predictor–outcome relationships—that is, effects of word properties in Language A on naming behavior in Language B.

Table 58
Regressions on Name Agreement Using All Four Length Metrics

Predictors	English	German	Spanish	Italian	Bulgarian	Hungarian
Goodness of depiction	+15.1‡	+6.9‡	+6.1‡	+4.5‡	+7.0‡	+7.0‡
Visual complexity	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Word frequency	+1.2‡	+3.2‡	+1.2‡	+2.4‡	+0.8†	n.s.
Initial frication	+0.7†	n.s.	n.s.	n.s.	n.s.	n.s.
Length in syllables	n.s.	+0.6*	n.s.	n.s.	+0.5*	n.s.
Syllable type frequency	n.s.	n.s.	n.s.	n.s.	n.s.	+0.8†
Length in characters	−0.7†	n.s.	n.s.	n.s.	−1.0†	n.s.
Word complexity	−0.4*	n.s.	−2.5‡	−1.1†	n.s.	−0.8†
Four length metrics entered together	(2.3‡)	(n.s.)	(4.1‡)	(3.0‡)	(n.s.)	(3.7‡)
Total variance	22.1‡	12.0‡	11.9‡	12.6‡	9.1‡	12.3‡

Note—Total variance and percentage of unique variance accounted for by each predictor on last step. Plus/minus indicates direction of contribution.   *$p < .1$.   †$p < .05$.   ‡$p < .01$.

Table 59
Regressions on Target Reaction Time Using Four Length Metrics

Predictors	English	German	Spanish	Italian	Bulgarian	Hungarian
Goodness of depiction	−31.5‡	−22.7‡	−20.1‡	−19.7‡	−24.6‡	−20.8‡
Visual complexity	n.s.	+0.6†	n.s.	n.s.	+0.7†	n.s.
Word frequency	−5.6‡	−7.0‡	−3.6‡	−7.9‡	−6.9‡	−3.4‡
Initial frication	n.s.	+0.5*	n.s.	n.s.	n.s.	n.s.
Length in syllables	n.s.	n.s.	n.s.	n.s.	−0.5†	n.s.
Syllable type frequency	n.s.	n.s.	−0.8†	−0.5†	n.s.	−0.6†
Length in characters	+0.6†	n.s.	n.s.	n.s.	+1.1‡	+0.5†
Word complexity	n.s.	n.s.	+1.1†	n.s.	n.s.	n.s.
Four length metrics entered together	(n.s.)	(n.s.)	(3.0‡)	(1.2‡)	(1.7‡)	(2.3‡)
Total variance	43.9‡	33.7‡	28.2‡	31.4‡	33.8‡	29.8‡

Note—Total variance and percentage of unique variance accounted for by each predictor on last step.
*$p < .1$.   †$p < .05$.   ‡$p < .01$.

To push this finding as far as we could, we developed novel summary measures of universal frequency and other-language frequency and compared the predictive value of these measures with own-language frequency. In all seven languages, other-language frequencies proved to be robust predictors of RT, even after own-language frequencies (as well as all the other word and picture properties) were controlled. When regressions were run in the opposite direction, own-language frequencies also contributed significant variance after other-language frequencies (and other word and picture properties) were entered into the equation. However, this residual effect of own-language frequency reached significance for name agreement only in English, Chinese, and German and reached significance for RTs only in English, Chinese, and Bulgarian. These are also the languages that have the "best" frequency measures (English, German, and Chinese are based on especially large written corpora; Bulgarian is based on subjective frequency ratings of the sort that sometimes outperform objective estimates of frequency; Balota, Pilotti, & Cortese, 2001). All of these results remained when analyses were repeated controlling for a rough (orthographic) estimate of cognate status. Hence, the other-language frequency effect is not coming entirely from words that *look the same* across these seven languages. Although we are not denying a role for word form frequency in picture naming, the robust effects of other-language frequency even after own-frequency is controlled suggests that frequency effects in picture naming are driven (at least in part) by deeper universal factors, such as conceptual accessibility or familiarity.

To test this hypothesis further, we asked whether frequency estimates (own language and other language) would facilitate RTs not only for the target name itself (the name used to derive own-language frequency), but also for the time required to produce alternative names (i.e., synonyms and morphological variants). If own- and other-language frequencies apply at the lexical level (lemma or word form), high-frequency target names ought to inhibit production of alternative names. Instead, we found exactly the opposite: Own- and other-language frequencies were negatively correlated with RTs not only for the target name itself, but also for the time required to produce synonyms or morphological variants. These results strongly suggest that some (although perhaps not all) of the frequency effects on picture-naming times comes from a conceptual level that is shared across languages, facilitating *any* of the names that speakers are able to find for that concept.

Similar cross-language effects were found for word length (i.e., length in Language A occasionally predicts naming behavior in Language B). However, these cross-language length results were smaller than the cross-language results reported above for own- and other-language frequency, and the universal effect of length disappears when the universal frequency factor is controlled. Hence we tentatively conclude that the cross-language effects of length are by-products of Zipf's law (i.e., the confound between frequency and length in every language).

Question III-3. What are the contributions of language-specific word structure properties to naming behavior? For example, in languages that tend to favor multisyllabic words, does syllable structure play a role on word retrieval that is not observed in languages in which short words prevail?

In contrast with the universal effects of goodness of depiction and frequency, word structure effects were small and variable over languages. But they do exist, and they can be seen for several different aspects of word structure. This includes some novel effects that have not been described in the literature on real-time language processing, such as the frequency of a particular syllable *template*. In short, word structure does matter, but it appears to matter in ways that vary from one language to another.

SUMMARY, CONCLUSIONS, AND FUTURE DIRECTIONS

We set out to investigate similarities and differences in timed picture naming across seven languages that vary along phonological, lexical, and grammatical dimensions that are known or suspected to play a role in lexical access. We have uncovered ample evidence for cross-linguistic differences, but the bulk of our findings suggest strong similarities in the factors that influence retrieval and production of words within this paradigm.

Although this study was not designed to decide among current theories of word production in general, the results have implications for those theories and suggest a number of directions for future research.

In Part I, we focused on similarities and differences in name agreement and RTs, measured in several different ways. The results included an across-the-board advantage for English, which is not surprising since our 520 picture stimuli were assembled in an English-speaking environment. But significant main effects of language remained on all variables when English was excluded from the analysis. In all of the tables comparing results for individual languages, we organized the data to reflect a hypothesized continuum of *language distance*: from English to German (two Germanic languages), to Spanish and Italian (two Romance languages), to Bulgarian (Slavic), and then to Hungarian and Chinese (the two non-Indo-European languages in our sample). However, we found relatively little evidence in favor of a language distance metric. Although Chinese and Bulgarian tended to show slower RTs and/or lower name agreement, none of the seven languages was an obvious outlier on any of our behavioral measures.

Correlations among the dependent variables were also quite similar within individual languages, including the finding that number of alternative names predicts naming latencies after percentage of agreement is controlled. This result is compatible with theories of naming that assume competition among alternative names prior to lexical selection.

All of the major dependent variables were significantly correlated across languages, suggesting that *hard* or *easy* items tend to be hard or easy for everyone. These conclusions were also supported by analyses using cross-language summary variables (e.g., average name agreement z scores, average RT z scores, or average number of alternative names). In fact, correlations using these summary variables were higher than the corresponding correlations within any individual languages, suggesting that we gained power and reliability by averaging our results. If relationships among dependent variables were driven in large measure by language-specific details, this should not have occurred.

In our view, the most significant result in Part I is the finding that cross-language correlations are much higher for RTs than for any other measure of naming behavior. This result underscores the value of a timed picture-naming paradigm, which appears to be sensitive to universal stages and/or processes that are not detected with off-line naming measures.

In Part II, we focused on similarities and differences in the target names (dominant response, or Lexical Code 1) that emerged in each language. Target word characteristics proved to be largely independent of picture characteristics (i.e., visual complexity and goodness of depiction), but there were enough small confounds to warrant inclusion of both word and picture characteristics in all the subsequent analyses. As was expected, we found substantial cross-language differences in word structure (length, frequency of different word templates, and initial frication). However, there were also significant correlations across languages on almost all measures of target word characteristics, including frequency, length, and initial frication. In other words, target names that are frequent, long, or fricative initial in one language tend to have the same characteristics in other languages as well. The small frication effect reflects the presence of cognates across all of these languages (including unrelated languages, such as English, Hungarian, and Chinese), and it is likely that similar correlations would appear if we had looked at another class of initial phonemes (initial frication was used because initial white noise can slow down detection of RTs by a voice key).

The length–frequency confound within languages reflects a well-known tendency for languages to assign shorter words to more frequent concepts (i.e., Zipf's law). Our cross-language results show that Zipf's law manifests itself both within and across languages, reflecting (we propose) cross-linguistic similarities in the frequency, familiarity, and/or accessibility of the concepts illustrated by our picture stimuli. We quantified this notion further by developing cross-language summary measures for both length and frequency, including universal frequency and universal length (based on the first principal component of factor analyses across all seven languages), together with some novel measures of other-language frequency and other-language length for each individual language.

In Part III, we examined the relationships between predictor and outcome variables, both within and across languages. Goodness-of-depiction ratings proved to be the best predictor of all naming measures in every language, although these ratings did work slightly better for English (the language in which they were compiled). In contrast, our objective measure of visual complexity had almost no effect on any aspect of naming behavior. Hence, we may conclude that goodness of depiction reflects cognitive factors in picture naming (and picture evaluation) that are largely independent of objective visual complexity—at least as we have measured it here (see Laws, Leeson, & Gale, 2002, for evidence that Euclidean overlap can slow down naming times under some conditions). Ratings had been obtained from a separate sample of English participants who were asked to rate how well each picture illustrates the empirically derived English target name. The fact that these ratings predicted naming behavior in seven different languages suggests that our English raters were evaluating how well each picture illustrates the intended concept, rather than the target word itself.

Word frequency was the second-best predictor of naming behavior, replicating a well-known result in the picture-naming literature (e.g., Oldfield & Wingfield, 1964, 1965). More surprising were the strong correlations that appeared across languages between frequency and naming behavior (e.g., Chinese frequencies predict naming behavior in Spanish). In fact, regression analyses showed that other-language frequencies made a sig-

nificant contribution to RTs when own-language frequencies were controlled in every language. A similar result was observed for within- versus cross-language effects of length, although this result appears to be a byproduct of the findings for frequency (evidence for the universality of Zipf's law).

These effects of cross-language frequency (and the shadow effects of cross-language length) may be the most important findings in this cross-linguistic study, because they force us to reconsider the locus of frequency effects in theories of picture naming. It is typically assumed that frequency effects are lexical in nature, reflecting the number of times that a speaker/listener is likely to have encountered a particular word form. However, in their review of the picture-naming literature, Johnson et al. (1996) warn us that this assumption is not always warranted, because word frequency may index the frequency, familiarity, and/or accessibility of the concept underlying that word. A number of studies have shown that frequency effects are *not* observed in picture processing *unless* lexical access is required (e.g., Caramazza, Costa, Miozzo, & Bi, 2001; Griffin, 2001; Kroll & Potter, 1984; Meyer, Sleiderink, & Levelt, 1998). This would suggest that frequency effects are not obligatory at the level of picture decoding per se. So where do these effects take place?

It has been suggested to us that our other-language frequency effects might constitute true word form frequency effects, via an indirect route. Frequency norms differ in size, coverage, reliability, and the context in which they were acquired. Even within a single language, one frequency measure sometimes "trumps" another, contributing significant variance to RTs after the first measure is controlled. From this point of view, it is possible that our other-language frequency measures are serving as an especially reliable index of word form frequency within each individual language. We did indeed find that *good* frequency measures based on large corpora (in English, German, and Chinese) and/or subjective ratings (Bulgarian) survived controls for other-language frequency, whereas weaker frequency measures based on smaller corpora (Italian and Hungarian) disappeared when other-language frequency was controlled. This suggests that the differential reliability of frequency measures is indeed playing a role. But other aspects of our findings suggest that something else is going on as well. Specifically, we found that target word frequency was associated not only with faster RTs for the target word itself, but also with faster RTs for alternative names (synonyms and morphological variants of the target word). This result suggests that the frequency effects are occurring not at the level of word selection (where a high-frequency target should prove to be a fierce competitor, slowing down RTs for alternative names), but at some point during the process by which a picture is recognized well enough to narrow down the lexical search. These conceptual effects also seem to percolate up to other levels of word form as well, reflected in the correlations among frequency and word structure variables (e.g., length, syllable type, complexity). At the very least, our results suggest that frequency effects are not restricted to the word form level. There may be at least two kinds of frequency variance in play: conceptual accessibility (which makes it easier to find any name for the picture in any language) and word form frequency (which is sensitive to variations in the reliability of the corpus from which the frequency measure is taken). We are currently trying to sort out these two possibilities by conducting cross-linguistic studies of word reading for comparison with the results presented here on picture naming, using the same target words. Insofar as word reading tends to be governed primarily by form-based factors, whereas picture naming is affected to a greater degree by conceptual factors, variations in the existence and magnitude of other-language frequency effects between reading and picture naming may provide useful information about the proposed mix of form-based and concept-based frequency effects.

Although we did find cross-linguistic effects of word structure, the existence and magnitude of these effects varied markedly from one language to another. Reliable and independent contributions of length and/or word complexity were observed in some languages (especially Italian and Spanish) but were weak or nonexistent in others (e.g., German). We also found independent effects of length in syllables and the relative frequency of different syllable templates (e.g., reliably faster RTs for disyllables in Chinese). Although all of these word structure effects were relatively small (as compared with the large and universal effects of goodness of depiction and frequency), they merit further investigation. Among other things, the small, language-specific effects observed in normal adults may be magnified in speaker/listeners who are working under a disadvantage, including young children in the course of language learning, older adults working with diminished processing resources, and individuals with congenital or acquired linguistic, cognitive, and/or perceptual–motor deficits.

On the basis of these results, we are expanding our inquiries in several new directions, including cross-language comparisons of action and object naming (e.g., Székely et al., 2003, for English), cross-language comparisons of word reading and auditory word repetition (using the same target words as those elicited in the present study), and developmental studies (e.g., D'Amico et al., 2001; Roe et al., 2000). We are also expanding our database to include new predictor variables. This will include a cross-linguistic exploration of AoA effects, using adult ratings of AoA, as well as objective measures taken from studies of early language development (see D'Amico et al., 2001; Iyer et al., 2001). Preliminary results suggest that cross-language AoA effects may be even larger than the cross-language frequency effects reported here, raising further concerns about conceptual versus lexical interpretations of predictor–outcome relationships. We are also expanding our database within each language to permit a more detailed exploration of word structure effects, using measures that are appropriate for each indi-

vidual language. This would include further studies of Chinese, in which we take advantage of the rich sublexical structure of compound words (which make up more than 80% of the words in this language). Within and across languages, the same materials are now being applied across clinical populations (including adult aphasics and children with language impairments), complemented by new initiatives in human brain imaging.

Finally, we are now in a better position to conduct the kinds of cross-linguistic studies of context effects that motivated us to obtain cross-linguistic norms in the first place. The main effects of language on naming variables that we reported here (e.g., Tables 3 and 4) suggest that some languages are "slower" than others. However, in view of the intense communicative pressures under which all languages evolve, it is quite unlikely that such a result will generalize when lexical access is studied in a discourse context. Every language has developed its own set of tradeoffs. For example, a historical move toward greater length is often compensated for by contextual cues that anticipate the point at which a word can be recognized (e.g., gender agreement cues that short-circuit the need for Italian listeners to wait for the gender markings that fall at the end of all nouns). The kinds of norms that we have developed and explored in the present study for lexical access out of context can be used to assess cross-linguistic variations in the contributions of semantic and grammatical cues to nature and timing of word recognition and production when these processes are embedded in a phrase, sentence, or discourse context.

REFERENCES

ABBATE, M. S., & LA CHAPELLE, N. B. (1984a). *Pictures, please! A language supplement.* Tucson, AZ: Communication Skill Builders.

ABBATE, M. S., & LA CHAPELLE, N. B. (1984b). *Pictures, please! An articulation supplement.* Tucson, AZ: Communication Skill Builders.

ALAMEDA, J. R., & CUETOS, F. (1995). *Diccionario de frecuencias de las unidades lingüísticas del castellano* [Frequency dictionary for lexical items in Castilian Spanish]. Oviedo: Servicio de Publicaciones de la Universidad de Oviedo. Available at http://www.swan.ac.uk/cals/calsres/vlibrary/jsg99a.htm.

ALCOCK, K. J., & NGOROSHO, D. (in press). Grammatical noun class agreement processing in Kiswahili. *Language & Speech.*

BAAYEN, R. H., PIEPENBROCK, R., & GULIKERS, L. (1995). *The CELEX Lexical Database (Release 2)* [CD-ROM]. Philadelphia: University of Pennsylvania, Linguistic Data Consortium [Distributor].

BACHOUD-LEVI, A., DUPOUX, E., COHEN, L., & MEHLER, J. (1998). Where is the length effect? A cross-linguistic study of speech production. *Journal of Memory & Language,* 39, 331-346.

BALOTA, D. A., & CHUMBLEY, J. I. (1985). The locus of word-frequency effects in the pronunciation task: Access and/or production? *Journal of Memory & Language,* 24, 89-106.

BALOTA, D. A., PILOTTI, M., & CORTESE, M. J. (2001). Subjective frequency estimates for 2,938 monosyllabic words. *Memory & Cognition,* 29, 639-647.

BATES, E., BURANI, C., D'AMICO, S., & BARCA, L. (2001). Word reading and picture naming in Italian. *Memory & Cognition,* 29, 986-999.

BATES, E., D'AMICO, S., JACOBSEN, T., SZÉKELY, A., ANDONOVA, E., DEVESCOVI, A., HERRON, D., LU, C.-C., PECHMANN, T., PLEH, C., WICHA, N. Y. Y., FEDERMEIER, K. D., GERDJIKOVA, I., GUTIERREZ, G., HUNG, D., HSU, J., IYER, G., KOHNERT, K., MEHOTCHEVA, T.,

OROZCO-FIGUEROA, A., TZENG, A., & TZENG, O. (in press). Timed picture naming in seven languages. *Psychonomic Bulletin & Review.*

BATES, E., DEVESCOVI, A., PIZZAMIGLIO, L., D'AMICO, S., & HERNANDEZ, A. (1995). Gender and lexical access in Italian. *Perception & Psychophysics,* 57, 847-862.

BATES, E., DEVESCOVI, A., & WULFECK, B. (2001). Psycholinguistics: A cross-language perspective. *Annual Review of Psychology,* 52, 369-398.

BATES, E., FEDERMEIER, K., HERRON, D., IYER, G., JACOBSEN, T., PECHMANN, T., D'AMICO, S., DEVESCOVI, A., WICHA, N., OROZCO-FIGUEROA, A., KOHNERT, K., GUTIERREZ, G., LU, C.-C., HUNG, D., HSU, J., TZENG, O., ANDONOVA, E., GERDJIKOVA, I., MEHOTCHEVA, T., SZÉKELY, A., & PLÉH, C. (2000). Introducing the CRL International Picture-Naming Project (CRL-IPNP). *Center for Research in Language Newsletter* (Vol. 12, No. 1). La Jolla: University of California, San Diego.

BATES, E., & MACWHINNEY, B. (1989). Functionalism and the competition model. In B. MacWhinney & E. Bates (Eds.), *The crosslinguistic study of sentence processing* (pp. 3-76). New York: Cambridge University Press.

BENTROVATO, S., DEVESCOVI, A., D'AMICO, S., & BATES, E. (1999). The effect of grammatical gender and semantic context on lexical access in Italian. *Journal of Psycholinguistic Research,* 28, 677-693.

CARAMAZZA, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology,* 14, 177-208.

CARAMAZZA, A., COSTA, A., MIOZZO, M., & BI, Y. C. (2001). The specific-word frequency effect: Implications for the representation of homophones in speech production. *Journal of Experimental Psychology: Learning, Memory, & Cognition,* 27, 1430-1450.

CARR, T. H., MCCAULEY, C., SPERBER, R. D., & PARMELEE, C. M. (1982). Words, pictures, and priming: On semantic activation, conscious identification, and the automaticity of information processing. *Journal of Experimental Psychology: Human Perception & Performance,* 8, 757-777.

CATTELL, J. (1886). The time to see and name objects. *Mind,* 11, 63-65.

CHEN, S. (1997). *Intra-lexical noun–verb dissociations: Evidence from Chinese aphasia.* Unpublished doctoral dissertation, University of Southern California.

CHEN, S., & BATES, E. (1998). The dissociation between nouns and verbs in Broca's and Wernicke's aphasia: Findings from Chinese. *Aphasiology,* 12, 5-36.

CHINESE KNOWLEDGE INFORMATION PROCESSING GROUP (1997). *Academia Sinica Balanced Corpus* (WWW Version 3.0). Taipei: Institute of Information Science, Academia Sinica. Available at http://www.sinica.edu.tw/ftms-bin/kiwi.sh.

COHEN, J., MACWHINNEY, B., FLATT, M., & PROVOST, J. (1993). PsyScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavior Research Methods, Instruments, & Computers,* 25, 257-271.

CUTTING, J. C., & FERREIRA, V. S. (1999). Semantic and phonological information flow in the production lexicon. *Journal of Experimental Psychology: Learning, Memory, & Cognition,* 25, 318-344.

CYCOWICZ, Y. M., FRIEDMAN, D., ROTHSTEIN, M., & SNODGRASS, J. G. (1997). Picture naming by young children: Norms for name agreement, familiarity, and visual complexity. *Journal of Experimental Child Psychology,* 65, 171-237.

DAMASIO, H., GRABOWSKI, T. J., TRANEL, D., PONTO, L. L. B., HICHWA, R. D., & DAMASIO, A. R. (2001). Neural correlates of naming actions and of naming spatial relations. *NeuroImage,* 13, 1053-1064.

D'AMICO, S., DEVESCOVI, A., & BATES, E. (2001). Picture naming and lexical access in Italian children and adults. *Journal of Cognition & Development,* 2, 71-105.

DAVIDOFF, J., & MASTERSON, J. (1996). The development of picture naming: Differences between nouns and verbs. *Journal of Neurolinguistics,* 9, 69-84.

DELL, G. S. (1990). Effects of frequency and vocabulary type on phonological speech errors. *Language & Cognitive Processes,* 5, 313-349.

DE MAURO, T., MANCINI, F., VEDOVELLI, M., & VOGHERA, M. (1993).

Lessico di frequenza dell'italiano parlato [Frequency lexicon of spoken Italian]. Milano: ETASLIBRI.

Dockrell, J., Messer, D., & George, R. (2001). Patterns of naming objects and actions in children with word-finding difficulties. *Language & Cognitive Processes*, **16**, 261-286.

Druks, J., & Shallice, T. (2000). Selective preservation of naming from description and the "restricted preverbal message." *Brain & Language*, **72**, 100-128.

Dunn, Lloyd M., & Dunn, Leota M. (1981). *Peabody Picture Vocabulary Test–Revised*. Circle Pines, MN: American Guidance Service.

Federmeier, K. D., & Bates, E. (1997). Contexts that pack a punch: Lexical class priming of picture naming. *Center for Research in Language Newsletter* (Vol. 11, No. 2). La Jolla: University of California, San Diego.

Füredi, M., & Kelemen, J. (1989). *A mai magyar nyelv szépprózai gyakorisgi szótra 1965–1977* [Prose frequency dictionary of contemporary Hungarian language 1965–1977]. Budapest: Akadémiai Kiadó.

Glaser, W. R. (1992). Picture naming. *Cognition*, **42**, 61-105.

Goodglass, H. (1993). *Understanding aphasia*. San Diego: Academic Press.

Griffin, Z. (2001). Gaze durations during speech reflect word selection and phonological encoding. *Cognition*, **82**, B1-B14.

Hernandez, A. E., Dapretto, M., Mazziotta, J., & Bookheimer, S. (2001). Language switching and language representation in Spanish–English bilinguals: An f MRI study. *NeuroImage*, **14**, 510-520.

Hernandez, A. E., Martinez, A., & Kohnert, K. (2000). In search of the language switch: An f MRI study of picture naming in Spanish–English bilinguals. *Brain & Language*, **73**, 421-431.

Hillert, D., & Bates, E. (1996). *Morphological constraints on lexical access: Gender priming in German* (Tech. Rep. 9601). La Jolla: University of California, San Diego, Center for Research in Language.

Humphreys, G. W., Riddoch, M. J., & Quinlan, P. T. (1988). Cascade processes in picture identification. *Cognitive Neuropsychology*, **5**, 67-103.

Iyer, G., Saccuman, C., Bates, E., & Wulfeck, B. (2001). A study of age-of-acquisition (AoA) ratings in adults. *Center for Research in Language Newsletter* (Vol. 13, No. 2). La Jolla: University of California, San Diego.

Jacobsen, T. (1999). Effects of grammatical gender on picture and word naming: Evidence from German. *Journal of Psycholinguistic Research*, **28**, 499-514.

Jescheniak, J.-D., & Levelt, W. J. M. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **20**, 824-843.

Johnson, C. J., Paivio, A., & Clark, J. M. (1996). Cognitive components of picture naming. *Psychological Bulletin*, **120**, 113-139.

Kaplan, E., Goodglass, H., & Weintraub, S. (1983). *Boston Naming Test*. Philadelphia: Lee & Febiger.

Kelly, M. (1992). Using sound to solve syntactic problems: The role of phonology in grammatical category assignments. *Psychological Review*, **99**, 349-364.

Kohnert, K. J. (2000). *Lexical skills in bilingual school-age children: Cross-sectional studies in Spanish and English*. Unpublished doctoral dissertation. University of California, San Diego and San Diego State University.

Kohnert, K. J., Bates, E., & Hernandez, A. E. (1999). Balancing bilinguals: Lexical-semantic production and cognitive processing in children learning Spanish and English. *Journal of Speech, Language, & Hearing Research*, **42**, 1400-1413.

Kohnert, K. J., Hernandez, A. E., & Bates, E. (1998). Bilingual performance on the Boston Naming Test: Preliminary norms in Spanish and English. *Brain & Language*, **65**, 422-440.

Kroll, J., & Potter, M. (1984). Recognizing words, pictures and concepts: A comparison of lexical, object, and reality decisions. *Journal of Verbal Learning & Verbal Behavior*, **23**, 39-66.

Lachman, R. (1973). Uncertainty effects on time to access the internal lexicon. *Journal of Experimental Psychology*, **99**, 199-208.

Lachman, R., Shaffer, J., & Hennrikus, D. (1974). Language and

cognition: Effects of stimulus codability, name-word frequency, and age of acquisition on lexical reaction time. *Journal of Verbal Learning & Verbal Behavior*, **13**, 613-625.

Laws, K., Leeson, V., & Gale, T. (2002). The effect of "masking" on picture naming. *Cortex*, **38**, 137-148.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral & Brain Sciences*, **22**, 1-38, 69-75.

Lu, C.-C., Bates, E., Hung, D., Tzeng, O., Hsu, J., Tsai, C.-H., & Roe, K. (2001). Syntactic priming of nouns and verbs in Chinese. *Language & Speech*, **44**, 437-471.

MacWhinney, B. (1987). The competition model. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 249-308). Hillsdale, NJ: Erlbaum.

Meyer, A., Sleiderink, A., & Levelt, W. J. M. (1998). Viewing and naming objects: Eye movements during noun phrase production. *Cognition*, **66**, B25-B33.

Murtha, S., Chertkow, H., Beauregard, M., & Evans, A. (1999). The neural substrate of picture naming. *Journal of Cognitive Neuroscience*, **11**, 399-423.

Nation, K., Marshall, C. M., & Snowling, M. J. (2001). Phonological and semantic contributions to children's picture-naming skill: Evidence from children with developmental reading disorders. *Language & Cognitive Processes*, **16**, 241-259.

Oldfield, R. C., & Wingfield, A. (1964). The time it takes to name an object. *Nature*, **202**, 1031-1032.

Oldfield, R. C., & Wingfield, A. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology*, **17**, 273-281.

Paivio, A. (1971). *Imagery and verbal processes*. New York: Holt, Rinehart, & Winston.

Roe, K., Jahn-Samilo, J., Juarez, L., Mickel, N., Royer, I., & Bates, E. (2000). Contextual effects on word production: A lifespan study. *Memory & Cognition*, **28**, 756-765.

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, **42**, 107-142.

Sanfeliu, M. C., & Fernandez, A. (1996). A set of 254 Snodgrass–Vanderwart pictures standardized for Spanish: Norms for name agreement, image agreement, familiarity, and visual complexity. *Behavior Research Methods, Instruments, & Computers*, **28**, 537-555.

Schmitt, B. M., Münte, T. F., & Kutas, M. (2000). Electrophysiological estimates of the time course of semantic and phonological encoding during picture naming. *Psychophysiology*, **3**, 473-484.

Schriefers, H., Meyer, A. S., & Levelt, W. J. M. (1990). Exploring the time course of lexical access in language production: Picture-word interference studies. *Journal of Memory & Language*, **29**, 86-102.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, familiarity and visual complexity. *Journal of Experimental Psychology: Human Learning & Memory*, **6**, 174-215.

Snodgrass, J. G., & Yuditsky, T. (1996). Naming times for the Snodgrass and Vanderwart pictures. *Behavior Research Methods, Instruments, & Computers*, **28**, 516-536.

Spieler, D. H., & Balota, D. A. (1997). Bringing computational models of word naming down to the item level. *Psychological Science*, **8**, 411-416.

Székely, A., & Bates, E. (2000). Objective visual complexity as a variable in studies of picture naming. *Center for Research in Language Newsletter* (Vol. 12, No. 2). La Jolla: University of California, San Diego.

Székely, A., D'Amico, S., Devescovi, A., Federmeier, K., Herron, D., Iyer, G., Jacobsen, T., Arévalo, A., Vargha, A., & Bates, E. (2003). *Timed action and object naming*. Manuscript submitted for publication.

Székely, A., D'Amico, S., Devescovi, A., Federmeier, K., Herron, D., Iyer, G., Jacobsen, T., & Bates, E. (in press). Timed picture naming: Extended norms and validation against previous studies. *Behavior Research Methods, Instruments, & Computers*.

van Turennout, M., Hagoort, P., & Brown, C. (1997). Electrophysiological evidence on the time course of semantic and phonological processes in speech production. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **23**, 787-806.

van Turennout, M., Hagoort, P., & Brown, C. (1998). Brain activity during speaking: From syntax to phonology in 40 milliseconds. *Science*, **280**, 572-574.

van Turennout, M., Hagoort, P., & Brown, C. (1999). The time course of grammatical and phonological processing during speaking: Evidence from event-related brain potentials. *Journal of Psycholinguistic Research*, **28**, 649-676.

Wicha, N. Y. Y., Bates, E., Moreno, E., & Kutas, M. (2000). Grammatical gender modulates semantic integration of a picture in a Spanish sentence. *Psychophysiology*, **37** (Suppl. 1). S104.

Wingfield, A. (1967). Perceptual and response hierarchies in object identification. *Acta Psychologica*, **26**, 216-226.

Wingfield, A. (1968). Effects of frequency on identification and naming of objects. *American Journal of Psychology*, **81**, 226-234.

Woehrmann, M. H., Hessenflow, A. N., Kettmann, D., & Yoshimune, P. H. (1994). *Image alchemy*. Freemont, CA: Handmade Software.

Zipf, G. K. (1965). *Human behavior and the principle of least effort: An introduction to human ecology*. New York: Hafner.