Behavior Research Methods, Instruments, & Computers
1996, 28 (4), 504–515

Statistical analysis of the bidirectional inconsistency of spelling and sound in French

JOHANNES C. ZIEGLER
Center for Research in Cognitive Neuroscience, Marseille, France

ARTHUR M. JACOBS
Center for Research in Cognitive Neuroscience, Marseille, France and Philipps-University, Marburg, Germany

and

GREGORY O. STONE
Arizona State University, Tempe, Arizona

Recent studies suggest that performance attendant on visual word perception is affected not only by the "traditional" feedforward inconsistency (spelling → phonology) but also by its feedback inconsistency (phonology → spelling). The present study presents a statistical analysis of the bidirectional inconsistency for all French monosyllabic words. We show that French is relatively consistent from spelling to phonology but highly inconsistent from phonology to spelling. Appendixes B and C list prior and conditional probabilities for all inconsistent mappings and thus provide a valuable tool for controlling, selecting, and constructing stimulus materials for psycholinguistic and neuropsychological research. Such large-scale statistical analyses about a language's structure are crucial for developing metrics of inconsistency, generating hypotheses for cross-linguistic research, and building computational models of reading.

When studying French as a second language, students often complain that French is unpredictable and ambiguous (Content, 1991). In a similar vein, French is often said to be highly inconsistent. *Inconsistency* typically refers to the ambiguity in the spelling-to-sound mapping. For instance, in English, the spelling body *-int* is inconsistent because it can be pronounced /aɪnt/ as in *pint* and /ɪnt/ as in *mint*. In contrast to French's reputation as an inconsistent language, recent estimates suggest that French orthography is more consistent than English orthography as far as the spelling-to-sound mapping is concerned (Content & Peereman, 1992). Gak (1976) claimed that 95% of words could be correctly deciphered on the basis of rules converting spelling to sound. The most precise estimate, taken from studies on automatic text-to-speech translation, lists no more than 204 exception words (Catach, 1984). Thus, French seems to be more consistent than intuition might suggest. To resolve this "inconsistency mystery" of French, detailed statistics about the bidirectional mapping of spelling and sound are needed.

For English, the mapping of spelling to sound has been studied extensively. Psycholinguists describe the spelling-to-sound structure in terms of rules that relate graphemes to phonemes (Venezky, 1970; Wijk, 1966). Neuropsychologists use inconsistency in the spelling-to-sound mappings as one of several tools for investigating different types of dyslexia (Patterson, Marshall, & M. Coltheart, 1985; Patterson & Morton, 1985; Plaut & Shallice, 1993). Educators and psychologists analyze the way in which children might learn to translate spelling to sound (Bosman & Van Orden, in press; V. Coltheart & Leahy, 1992; Treiman, Mullennix, Bijeljac-Babic, & Richmond-Welty, 1995; Waters, Seidenberg, & Bruck, 1984; Wimmer & Goswami, 1994). Cognitive psychologists propose models and theories to account for the various facets of inconsistency observed in naming and lexical decision tasks (M. Coltheart, 1978; Forster & Chambers, 1973; Frederiksen & Kroll, 1976; Glushko, 1979; Taraban & McClelland, 1987; Van Orden & Goldinger, 1994; Waters & Seidenberg, 1985). Finally, proponents of computational models are challenged to implement processes in simulation models that capture human performance for inconsistent words (Brown, 1987; M. Coltheart, Curtis, Atkins, & Haller, 1993; M. Coltheart & Rastle, 1994; Norris, 1994; Seidenberg & McClelland, 1989; Van Orden, Bosman, Goldinger, & Farrar, in press).

Not surprisingly, detailed statistical descriptions of the spelling-to-sound relation in English are available and

The research reported in this article was supported by a German Academic Exchange Service Grant (DAAD-Doktorandenstipendium aus Mitteln des zweiten Hochschulsonderprogramms) awarded to J. C. Ziegler. We thank M. Coltheart and D. Massaro for stimulating discussions on this and related topics and R. S. Berndt, S. Ransdell, and an anonymous reviewer for valuable feedback on an earlier version of this article. We are grateful to A. Content, M. Montant, and A. Rey for their help with the subtleties of French orthography and phonology. Correspondence should be addressed to J. C. Ziegler, Centre de Recherche en Neurosciences Cognitives, Centre National de la Recherche Scientifique, 31, Chemin Joseph Aiguier, 13402 Marseille Cedex 20, France (e-mail: ziegler@lnf.cnrs-mrs.fr).

provide valuable research tools for planning experiments and constructing stimulus materials (Berndt, D'Autrechy, & Reggia, 1994; Berndt, Reggia, & Mitchum, 1987; P. R. Hanna, J. S. Hanna, Hodges, & Rudorf, 1966). Detailed analyses of a language's structure seem crucial for developing and testing models of reading in which the processing of one linguistic item is influenced by the entire set of items the model knows, as in current connectionist models of reading (Grainger & Jacobs, 1996; Jacobs & Grainger, 1992, 1994; McClelland & Rumelhart, 1981; Stone & Van Orden, 1994; Ziegler, Rey, & Jacobs, 1995; see Treiman et al., 1995, for a similar argument).

To determine whether a monosyllabic word is inconsistent, it is typically broken down into its onset and spelling body. The *onset* is the initial sequence of consonants, and the spelling *body* (or rime) is everything following it (see, e.g., Patterson & Morton, 1985; Treiman et al., 1995). For example, *pint* can be divided into the onset *p-* and the spelling body *-int*. Words are traditionally classified as *inconsistent* if their spelling body maps into more than one pronunciation (e.g., *-int* in *pint* vs. *mint*). They are traditionally called *consistent* if their spelling body has only one possible pronunciation (e.g., *-uck* in *duck*, *luck*).

In numerous studies, the inconsistency of the spelling-to-sound mapping has been shown to affect performance in lexical decision and naming tasks (Andrews, 1982; V. Coltheart & Leahy, 1992; Content, 1991; Content & Peereman, 1992; Glushko, 1979; Jared, McRae, & Seidenberg, 1990; Seidenberg, Waters, Barnes, & Tanenhaus, 1984; Taraban & McClelland, 1987; Waters & Seidenberg, 1985). In the naming task, it takes longer to read aloud inconsistent words, such as *pint*, than it does to read consistent words such as *duck* (*-int* in *pint* may be pronounced as in *mint*; *-uck* in *duck* is pronounced only as in *duck*). Occasionally, skilled readers make regularization errors. They may incorrectly pronounce *pint* to rhyme with *mint*. Those regularization errors are characteristic of surface-dyslexic patients. In the lexical decision task, participants produce slower responses and more errors to inconsistent items than to consistent items. These consistency effects are stronger for low-frequency than for high-frequency inconsistent items and are often statistically reliable only for low-frequency words (but see Jared, 1995, for consistency effects for high-frequency words). To summarize, these spelling → phonology consistency effects raise the question of how reading is affected when a spelling has more than one pronunciation.

A New Perspective

Until recently, all discussion of consistency effects has concerned a classic "feedforward" spelling → phonology effect. However, Stone, Vanhoy, and Van Orden (in press) challenged this "one-way-inconsistency" perspective. They demonstrated that visual word perception is not only influenced by "traditional" spelling-to-phonology inconsistency but also by phonology-to-spelling inconsistency (i.e., when a phonologic body maps into more than one spelling). According to Stone et al.'s terminology, words are *feedforward inconsistent* if a word's spelling body has more than one possible pronunciation, such as *-int* in *pint* and *mint*. Words are *feedback inconsistent* if a word's phonologic body has more than one possible spelling, such as /-ip/ in *deep* and *heap*. In fact, Stone et al. found that lexical decision latencies to words that were traditionally labeled as consistent were longer if they were feedback inconsistent than if they were feedback consistent.

In a different line of research, Ziegler and Jacobs (1995; see also Ziegler, Van Orden, & Jacobs, in press) recently demonstrated that feedback inconsistency also affects performance in simple graphemic tasks. In a letter search task, they presented pseudohomophones, like *brane*. Pseudohomophones and homophones are, by definition, feedback inconsistent because their phonology can be spelled in more than one way. The authors found that letter detection performance was worse (longer reaction times [RTs] and more errors) for feedback-inconsistent letter strings (i.e., pseudohomophones) than for feedback-consistent spelling controls.

In previous studies of visual word recognition, inconsistency in terms of sound to spelling (feedback inconsistency) has been a neglected source of information. However, a statistical analysis of all English monosyllabic words showed that feedback inconsistency is common. Stone et al. (in press) calculated that 76.6% of the monosyllabic words taken from Kučera and Francis (1967) were feedback inconsistent and 72.4% of the word occurrences were feedback inconsistent. On the basis of these results, they concluded that the ignored factor of feedback inconsistency could have been responsible for small and unreliable consistency effects in previous studies, because investigators might have failed to control for feedback inconsistency in their consistent control items (against which the processing costs of inconsistent items were tested)—that is, whether their consistent words contained phonologic bodies that could be spelled in more than one way. To summarize, recent studies have shown that it matters in visual word perception when a pronunciation has more than one spelling. Furthermore, feedback inconsistency is rather common in English. In the present study, we look at feedforward and feedback inconsistency in French.

The Present Study

The study and statistical analysis by Stone et al. (in press) provides a key to understanding the aforementioned "inconsistency mystery" of French. French might be relatively consistent from spelling to phonology, but much less consistent from phonology to spelling. In the present study, we investigate this hypothesis. We provide a statistical database giving the degree of inconsistency in both directions: from spelling to phonology (feedforward) and from phonology to spelling (feedback). Furthermore, we compare inconsistency in English to inconsistency in French. In doing so, we hope to stimulate cross-linguistic research by generating hypotheses about