Journal of Phonetics 53 (2015) 153–176

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/phonetics

Bridging phonological system and lexicon: Insights from a corpus study of functional load

Yoon Mi Oh*, Christophe Coupé, Egidio Marsico, François Pellegrino

Laboratoire Dynamique du Langage, UMR5596 Université de Lyon and CNRS, France

ARTICLE INFO

Article history:
Received 13 November 2014
Received in revised form
27 July 2015
Accepted 25 August 2015
Available online 23 October 2015

Keywords:
Functional load
Phonological systems
Mental lexicon
Cross-linguistic
Consonants and vowels
Phonological contrast

ABSTRACT

In this paper, we propose a functional and cross-language perspective on the organization of phonological systems based on the notion of functional load (FL). Using large corpora, we quantitatively characterize the relationships between phonological components (segments, stress and tones) by estimating their role at the lexical level. In a first analysis, we examine the relative contribution of each phonological subsystem to the pool of lexical distinctions and compare the results between two tonal (Cantonese and Mandarin) and seven non-tonal languages (English, French, German, Italian, Japanese, Korean, and Swahili). The equal weight of vowels and tones in lexical distinction is confirmed as well as the phenomenon of consonantal bias – advocated in several psycholinguistic studies – in five languages (English, French, German, Italian, and Swahili), with various corpus configurations in order to assess the influence of morphology and usage frequency. Our results reflect a strong preference toward consonant-based distinctions rather than vowel-based distinctions in a reduced (lemmatized) configuration of the lexicon. This preference is nevertheless modulated when inflectional morphology and usage frequency were considered. A second analysis consists in a cross-language comparison of the internal FL distribution within vocalic and consonantal subsystems in nine languages. We observe uneven FL distributions with only a few salient high-FL contrasts.

Shared trends in terms of the mostly employed phonological features are also revealed but a few language-specific patterns are also present. These results are discussed in terms of organization and processing of the mental lexicon.

1. Introduction

1.1. The concept of functional load

As stated by Hockett, “The function of a phonemic system is to keep the utterances of a language apart” (Hockett, 1966:1). Phonemes are thus considered the elementary bricks on which contrasts between words are built. The most obvious procedure to identify them is by listing minimal pairs (when they exist): two sound sequences associated with two different meanings and differing by only one element. The set of such ‘distinctive’ elements constitutes the phonemic system of a particular language. For decades, studying phoneme inventories has been the gateway for understanding how languages work. This traditional approach to phonemes and relations between them has yielded highly significant insights into the organization of phonological systems (Crothers, 1978; Hall, 2011; Hyman, 2008; Liljencrants & Lindblom, 1972; Lindblom, 1986; Lindblom & Maddieson, 1988; Maddieson, 1984; Marsico, Maddieson, Coupé, & Pellegrino, 2003; Schwartz, Boë, Vallée, & Abry, 1997; Vallée, 1994). However, a side-effect of this paradigm is that, because all phonemes in an inventory are given the same importance, disregarding their frequency and their role in contrasts,[1] certain key phenomena remain underappreciated. To illustrate, consider asking a British English (RP: Received Pronunciation)

* Corresponding author.
E-mail addresses: yoon-mi.oh@univ-lyon2.fr (Y.M. Oh), christophe.coupe@cnrs.fr (C. Coupé), egidio.marsico@cnrs.fr (E. Marsico), francois.pellegrino@univ-lyon2.fr (F. Pellegrino).

[1] Even if vowels and consonants (as well as their natural subsets: stops, fricatives, etc.) are not considered identical, in terms of production (Ladefoged & Maddieson, 1996), acoustics (Fogerty & Humes, 2012; Ladefoged, 2001; Stevens, 2002; among others), and perception (Fry, Abramson, Eimas, and Liberman 1962; Kronrod, Coppess, & Feldman, 2012; Liberman, Harris, Hoffman, and Griffith, 1957). These differences have recently been mirrored by neurophysiological findings (Caramazza, Chialant, Capasso, & Miceli 2000; Mesgarani, Cheung, Johnson, & Chang 2014; Obleser, Leaver, VanMeter, & Rauschecker 2010; Scharinger, Idsardi, & Poe, 2011). Vowels and consonants are not identical in terms of functional role either (Nespor, Peña, & Mehler, 2003; New, Araújo, & Nazzi, 2008; Toro, Nespor, Mehler, & Bonatti 2008), should it be defined by usage frequency or FL, for instance.

0095-4470/$ - see front matter © 2015 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.wocn.2015.08.003

speaker to provide an example of a minimal pair based on a consonantal contrast. Her answer is likely to include word pairs that exhibit a "high frequency" contrast such as /t-d/ (as in "*tip*" vs. "*dip*"), as opposed to word pairs that exhibit a "low frequency" contrast such as /ʒ-v/, (as in "*closure*" /ˈkləʊʒə/ vs. "*clover*" /ˈkləʊvə/). The point is that some phonemic contrasts in English, differentiate hundreds of word pairs (e.g. /t-d/) while others may only be involved in a handful of word pairs (e.g. /ʒ-v/). This fact accords with Hockett's addendum to his characterization of the functional role of phonemes: i.e. that "Some contrasts between the phonemes in a system apparently do more [keeping apart of words] than others" (Hockett, 1966:1). Moreover, this observation appears to hold true for other languages as well, with the work done by particular contrasts potentially varying across languages. Indeed, the Prague School thought that specific contrasts may differ from one language to another and that this "rendement fonctionnel" or "charge fonctionnelle" (Functional Load, henceforth FL) should be taken into consideration when reasoning about phonological systems (Cercle Linguistique de Prague, 1931; Jakobson, 1931).

1.2. Some landmarks on functional load

Despite a general agreement on what it covers, it should be noted that the concept of FL has often been considered in an impressionistic way (for a review, see Surendran & Niyogi, 2003). As a consequence, FL is generally described by circumlocutions and no precise theoretical definition exists, beyond general statements such as "The term FUNCTIONAL LOAD is customarily used in linguistics to describe the extent and degree of contrast between linguistic units, usually phonemes" (King, 1967). To be fair, one should also note that formal mathematical definitions arose as early as the mid-fifties (Hockett, 1955) and provided enough ground to address FL-related issues. Before this quantitative characterization, advocates of FL heavily relied on intuitions and extensions of the notion of phonological contrast. As stated in the previous section, phonological contrast and opposition were central concepts within the Prague School. Trubetzkoy later mentioned that an "economical" language would very often distinguish words by only one phoneme while "prodigal" languages would make usage of several phonological elements to keep words distinct (Trubetzkoy, 1939:240). Kučera (1963) compared phonemic and syllabic inventory entropies, as well as some derived FL measures, in Russian and Czech. Yet, references to FL have remained sporadic for decades, probably because of the difficulty to process large corpora, which were moreover hardly available. This state lasted until Surendran and Niyogi breathed new life into the concept at the beginning of this century. They compared FL of tones, stress, phonemes and phonetic features in four languages (Dutch, English, German, and Mandarin) and highlighted the importance of the tonal system in Mandarin (Surendran & Niyogi, 2003). This result was confirmed in a follow-up study (Surendran & Levow, 2004) and recently extended to Cantonese (Oh, Pellegrino, Coupé, & Marsico, 2013). Oh and colleagues also compared the relative functional weight of consonantal, vocalic (and tonal, if any) systems in five languages (Cantonese, English, Japanese, Korean, and Mandarin). Their results suggest that the distributions of FL in a phonological system are very uneven, with only a few prominent contrasts. These differences in relative prominence may be useful to take into consideration for foreign language acquisition (following Brown, 1988; Munro & Derwing, 2006).

Besides typology-oriented studies, the main topic for which FL was considered relevant was historical linguistics. Upon its inception, Martinet promoted the notion of FL, suggesting that it may play a role in language change (Martinet, 1938, 1955). According to his hypothesis, also adopted later by Hockett (1966), phonemes involved in high-FL contrasts would be less prone to merging than those involved in low-FL contrasts. Corpus-based studies have failed to confirm this hypothesis for decades (King, 1967; Surendran & Niyogi, 2003; Surendran & Niyogi, 2006), but a recent cross-language study brought some support to it (Wedel, Kaplan & Jackson, 2013). Such conflicting results may be due to differences in corpora or to the small number of sound changes considered so far. It is also possible that, even if FL plays a role in phonetic change, its magnitude is limited, for example with regard to social factors (Labov, 2001). As a consequence, even if FL does determine a pool of potential changes, their actual implementation in a language or a dialect probably depends on further aspects.

From a different angle, the availability of corpora in the field of child language acquisition also stimulated interest in the notion of FL. Its impact on the order of phoneme acquisition by children was demonstrated (Pye, Ingram, & List, 1987; Van Severen et al., 2012), in conjunction with language-specific properties (Stokes & Surendran, 2005). Again, FL is not the only factor at play in the course of phonological acquisition, but converging cues indicate that the phonemes involved in high-FL oppositions have a tendency to be acquired earlier than the others (Van Severen et al., 2012). Stokes and Surendran (2005) showed nevertheless that the effect of FL should be considered with caution since FL was not a significant predictor of consonant order of acquisition in Cantonese-speaking children, in contrast with what they observed in English-speaking children (Stokes & Surendran, 2005).

This re-emergence of the concept of FL can be seen as part of a general movement for promoting statistical and information-theoretic quantitative approaches (see Goldsmith, 2000). Today for instance, the relevance of usage frequency is well acknowledged, and many studies in psycholinguistics, phonology, and phonetics have proven that it significantly impacts cognitive processes, such as access to mental representations (Bybee, 2003; Cholin, Levelt, & Schiller, 2006; Jescheniak & Levelt, 1994; Johnson, 1996; Levelt, Roelofs, & Meyer, 1999; Pierrehumbert, 2001; Schilling, Rayner, & Chumbley, 1998; Walsh, Möbius, Wade, & Schütze, 2010). It has nevertheless been less often mentioned in the study of phonological systems per se. However, we think that taking this functional approach into consideration can notably change our vision of phonological systems and can enrich our knowledge of speech cognitive processing. The goal of this paper is consequently to shed new light on phonological systems from the perspective of FL. The emphasis is placed on both their internal functional organization and their importance in building the lexicon. Results are then discussed on communicative and cognitive grounds, in connection with the main focus of this Special Issue.

For almost one century, FL has thus been suggested as a factor involved in the *acquisition* and the *evolution* of phonological units and systems as well as a *systemic* property rooted in lexical strategies. These three dimensions have in common the fact that they