Functional load revisited
Reinterpreting the findings of ‘lingua franca’ intelligibility studies

Andrew Sewell
Lingnan University

The concept of functional load has been invoked for nearly a century in both research and pedagogy relating to pronunciation. However, it often suffers from a certain vagueness of definition. This article revisits the concept of functional load, outlining its origins and placing it within functionalist approaches to language and communication. The elaborated concept is then applied to the findings of lingua franca intelligibility studies, and it is argued that functional factors have explanatory potential. The application of functional load thus serves as a reinterpretation of the findings of lingua franca intelligibility studies by providing additional theoretical background. This reinterpretation suggests that although functional load is not a universal property of abstract language systems, certain features (such as consonants in general, and initial segments in particular) are widely relied upon in both lingua franca and non-lingua franca communication.

1. Introduction

The outlook for pronunciation research in the 21st century includes promising opportunities and interesting challenges, as outlined by Munro and Derwing (2015). There are also some pitfalls to be avoided. One of these is the tendency for ‘old ideas’ to be presented as ‘new revelations’, and the ‘relatively recent lingua franca view’ is cited as an example (Munro & Derwing, 2015: 15). The lingua franca view focuses on interactions between non-native speakers, and one of its main indications is that these speakers ‘need not emulate particular native pronunciations’ (Munro & Derwing, ibid.). But this is by no means a new idea (see, e.g., Abercrombie, 1949; Gimson, 1962; Sweet, 1964 [1899]). The ‘lingua franca view’
provokes a strong sense of *déjà vu* in some quarters, and as such it appears to be little more than a restatement of the long-standing 'intelligibility principle' (Levis, 2005; see also Levis, 2016).

Studies of intelligibility in lingua franca interactions have played a major role in the development of the lingua franca view. However, the small sample sizes in much of this research have drawn criticism (e.g., Munro & Derwing 2015: 34). The frequently-cited Lingua Franca Core (LFC) research was derived from only 27 pronunciation-related instances of misunderstanding (Jenkins, 2000: 85). The methodological assumptions of such research have also been questioned. One might therefore wonder whether it is worth giving continued attention to either the lingua franca approach to intelligibility or its implications for pronunciation pedagogy.

In general, despite the difficulties and the need for complementary approaches, there are strong arguments in its favour. One cannot travel far in today’s world without encountering ‘lingua franca’ communication, in fields as diverse as English-medium tertiary education, sport and international aviation. In fact, given the diversity of interactional patterns it is certainly possible to argue that we are all lingua franca speakers, regardless of our language backgrounds. If there are plausible explanations for instances of lingua franca *mis*communication then it is worth considering them, not only for their practical relevance but also for what they can tell us about communication in general. However, in order to do this and address the criticisms outlined above, two things are needed: more data relating to instances of misunderstanding, and a stronger theoretical foundation from which to consider their possible causes.

Accordingly, in this review article I compare the findings of the LFC with those of a more recent lingua franca intelligibility study (Deterding, 2013) that draws on a larger database. This revisitation will also involve a critical perspective on the notion of causation in such research. I then reinterpret the findings of these studies in the light of a somewhat neglected theoretical concept, that of *functional load* (see, e.g., Brown, 1988; Catford, 1987, 1988; Gilner & Morales, 2010; Munro & Derwing, 2006; Rischel, 2007; Surendran & Niyogi, 2003, 2006). It is notable that despite the frequent calls made for further research into this important concept – a call repeated in the prospectus of Munro and Derwing (2015: 18) – little attention has been paid to it, and few attempts have been made to apply the concept to the findings of the LFC and similar research.

One of the aims of this article is thus to take a new look at the concept of functional load, evaluating its usefulness in explaining previous studies or informing future ones. To do this, it provides a synthesis of perspectives on the concept, including the linguistic discussions of the Prague School, the information-theoretic approach, and evidence from psycholinguistic studies. The concept is elaborated and linked with functionalist approaches to language and communication (e.g.,
Functional load revisited

Bybee, 2001; Bybee & Hopper, 2001), to increase its theoretical coherence. The other main aim of the article is to reinterpret the findings of lingua franca intelligibility studies in the light of functional load; the elaborated concept is therefore applied to the findings of Jenkins (2000) and Deterding (2013). It is argued that this version of functional load has some explanatory value and can be used to make some useful generalizations, keeping certain caveats in mind. The article concludes by reviewing the main implications, for research and pedagogy, of the concept of functional load.

2. Lingu franca intelligibility studies: Methods and findings

The research interest in ‘lingua franca’ interactions between non-native speakers (NNSs) of English reflects both the claimed numerical dominance of NNSs in international communication and the relative paucity of intelligibility studies in this domain (see Pickering, 2006).1 A notable subtype of these lingua franca studies involves the creation and analysis of corpora of misunderstandings. In this section I will review the methods and findings of two corpus-based, lingua franca intelligibility studies: Jenkins (2000), which is based on data from several studies carried out by the researcher, and Deterding (2013), which analysed interactions among nine participants from different countries and regions in south-east Asia (Brunei, Malaysia, Indonesia and Laos), east Asia (Hong Kong, Taiwan, China and Japan) and Africa (Nigeria).

2.1 Methods

The first stage of data analysis in these corpus-based studies is the identification of instances of misunderstanding, which are subdivided into categories according to their major cause. In one of the studies reported in Jenkins (2000), 40 instances of misunderstanding were found to contain 27 pronunciation-related cases (68%); Deterding (2013) collected 183 tokens of misunderstanding, 158 (86.3%) of which were pronunciation-related. These figures are interesting in themselves, as they suggest that pronunciation may be the major cause of misunderstanding in lingua franca interactions (see Jenkins, 2000: 83).

Numerous methodological questions arise, however, such as whether these instances represent problems of intelligibility (as opposed to some other construct), and how these instances of misunderstanding can be identified. The answers

1. It is worth noting that many earlier studies of non-native interaction, for example those taking place in ESL classroom settings, could be reframed and re-examined as ‘lingua franca’ studies.
depend, at a foundational level, on how ‘intelligibility’ is defined. Deterding (2013) broadly adopts the definition of Smith and Nelson (1985): intelligibility is concerned with the recognition of words and utterances (see also Nelson, 2008, 2011). For Deterding (2013: 10), an instance of misunderstanding occurs when ‘there are some key words that the listener cannot identify or does not understand, even if at the wider level they may be able to follow the gist of the utterance quite successfully’. In this approach, intelligibility is thus largely separated from comprehensibility (‘the capacity to assign meanings to what we hear or read’; Nelson, 2011: 34). It is worth noting that these definitions of intelligibility and comprehensibility are quite different from those of Munro and Derwing (e.g., Munro, Derwing & Morton, 2006), who differentiate them on the basis of actual understanding versus perceived ease of understanding.

In Jenkins’s LFC research, instances of misunderstanding were identified and noted by the researcher as they occurred. One limitation of this approach is that the occurrence of ‘misunderstanding’ is not always apparent to observers (as noted by Deterding, 2013: 13). Deterding’s own study took a broader view of misunderstanding, and addressed the identification problem by involving participants in the process of transcribing conversations and identifying mismatches. An illustration of this process is provided by Deterding (2013: 25–26): a Bruneian listener transcribed part of a Nigerian speaker’s utterance as *out kick*, and subsequent explanation from the speaker established his intention as *hot cake*. Deterding’s methodology is therefore capable of addressing ‘actual understanding’ (part of the definition of intelligibility adopted by Munro & Derwing, 2015: 14). It has the important advantage of involving participants in the identification of ‘misunderstanding’, rather than privileging the analyst’s perspective (see also Zielinski, 2008 for an example of this approach with native listeners).

With 158 pronunciation-related tokens of misunderstanding, Deterding (2013) also addresses one of the main limitations of the corpus-based approach, namely the small data sets used in many studies (see Deterding, 2013: 26, for commentary). However, it has to be pointed out that Deterding’s data set is not quite as varied as the headline figures suggest. Although there are nine speakers from different countries, some speakers were misunderstood more often and by more listeners, and are thus associated with more than their fair share of tokens of misunderstanding. The Hong Kong speaker is analysed as contributing 68 tokens (37%) to the overall corpus, and the Indonesian speaker is linked with 46 (25%). In fact, 157 of the 183 tokens of misunderstanding (86%) occurred in conversations with the four male speakers (from Hong Kong, Indonesia, Nigeria and Laos), leaving
the five female speakers with only 14%. Whatever the causes of this imbalance, generalizations to other populations need to be made with caution.\footnote{A reviewer commented that this analysis suggests it is speakers who are responsible for (or who ‘contribute’) instances of misunderstanding, whereas intelligibility problems are always a result of interacting speaker and listener factors. This is accepted, but in Deterding’s study the analysis of tokens by speaker is to some extent justified by the methodology: the same speakers were heard on different occasions by different listeners, and plausible pronunciation-related causes of misunderstanding were identified.}

In general methodological terms, the \textit{a posteriori} approach of the LFC and other corpus-based studies of intelligibility has long been regarded with skepticism in applied linguistics (Murray Munro, personal communication). The central problem is that although identifying instances of intelligibility problems is relatively straightforward, identifying their causes is difficult because there are multiple interacting factors at work. One facet of this co-occurrence problem is that it is hard to separate speaker factors (such as pronunciation) from listener factors (such as familiarity with words). For example, when meal plan (from a Hong Kong speaker) was misunderstood as view pen (by a listener from Taiwan), Deterding (2013: 66) notes that the term meal plan was ‘not familiar’ to the Taiwanese listener. The misunderstanding was thus, arguably, a result of both the speaker’s pronunciation and the listener’s unfamiliarity with the collocation of these words. The notion of ‘familiarity with words’ remains difficult to operationalize, but assessing it is facilitated if participants are able to comment on their conversations, as in Deterding’s study (see also Munro & Derwing, 2009, for another example of the ‘introspective’ approach).

Another facet of the co-occurrence problem is deciding which pronunciation features caused the misunderstanding. The meal plan/view pen incident was categorised under both the ‘vowel’ and ‘initial cluster’ headings; the TRAP vowel in plan was closer to DRESS, and [l] was not pronounced in the initial cluster (Deterding, 2013: 66). A third factor is also considered by Deterding (2013: 96), namely the L-vocalisation in meal. With the \textit{a posteriori} approach it becomes difficult to decide which factors are most important, except through frequency rank- ings and comparisons with other studies.

Even if frequency rankings are produced, however, they tend to mask a more fundamental problem with the \textit{a posteriori} approach. This is that the \textit{absence} of a particular type of feature in such analyses is ‘largely meaningless’ (Murray Munro, personal communication; see also Schachter, 1974). For example, both Jenkins (2000) and Deterding (2013) found that consonantal modifications were the major cause of intelligibility problems (e.g., ‘the greatest impact on intelligibility comes from consonants, which is consistent with the LFC proposals’; Deterding,
2013: 91). But although vowel modifications were relatively absent as the cause of intelligibility problems, this does not necessarily mean that vowels are less important, for the simple reason that we have no easy way of knowing the frequency and extent of vowel modifications in the dataset.

Because of this, one can argue that there are two completely different scenarios, both of which are capable of explaining the smaller number of vowel-related problems. It may be that the speakers made many vowel modifications, in which case it would be reasonable to conclude that the exact realization of vowels is not particularly important. On the other hand, it may be that speakers made relatively few significant vowel modifications, which would suggest that vowels do matter—so much so that the speakers had already learned to produce vowels within certain limits of variation.

To put it another way, absence of evidence is not evidence of absence. The relative absence of vowel-related instances of misunderstanding does not necessarily mean that vowels are unimportant, upholding the validity of Munro’s criticism. But despite these problems, it has to be acknowledged that dealing with multiple causation is inevitable in the analysis of naturally-occurring conversation; it is the price paid for greater realism, and represents a familiar trade-off between internal and external validity. The lingua franca approach has significant contributions to make to an overall understanding of intelligibility, especially if its findings are viewed alongside those of other approaches that seek greater control over speaker and listener variables.

In defense of the *a posteriori* approach, one could make parallels with other disciplines that seek to explain complex social phenomena. In the social sciences more generally, Lawson (1997) identifies two ‘distinct but complementary sets of activities’ (see Elder-Vass, 2012: 18). One of these is called retroduction, and involves identifying single causal powers and explaining the mechanisms that produce them. In intelligibility research, the studies of primary phrase stress and lexical stress carried out by Hahn (2004) and Field (2005) respectively are predominantly retroductive in orientation, as they sought to control and compare the effects of specific variables.

The other, complementary process is called retrodiction, and it is concerned with events, rather than single causal powers. It involves identifying the most significant members of the set of powers that interacted to cause the event concerned, and how they interacted with each other (Elder-Vass, 2012: 19). As Elder-Vass notes, an empirical science must be able to engage in both. Corpus-based studies of natural conversation therefore have their place, alongside experimental manipulations. However, it is noticeable that while both Jenkins (2000) and Deterding (2013) identify causal powers in the form of lists of features, they provide few explanations for how these features serve to maintain or disrupt intelligibility. A
possible reason for this has been the lack of a theoretical framework that can help to explain both causal powers and particular communicative events. Later in this article I will discuss the concept of functional load, which I argue can contribute to such a framework, but first it is necessary to briefly review and summarise the findings of these lingua franca intelligibility studies.

2.2 Findings

The Lingua Franca Core (LFC) of Jenkins (2000; 2002) is described as a list of features whose presence acted as ‘safeguards of mutual intelligibility’ in lingua franca interactions (Jenkins, 2002: 96). In other words, they are features whose presence should be encouraged and whose absence should be discouraged. The main constituents of the LFC can be grouped under four headings:

1) Consonants: most of the consonants in so-called ‘standard’ inventories such as RP and GA must be retained. Substitutions of the dental fricatives and alternative realisations of postvocalic /l/ (including L-vocalisation) are acceptable. The general principle, however, is that approximations should not ‘overlap onto another phoneme’ (Jenkins, 2002: 96).

2) Consonant clusters: initial clusters are important for intelligibility and cannot be simplified. Final clusters can be simplified, but only in ways that resemble native-speaker patterns (Jenkins, 2002: 97).

3) Vowels: contrasts based on long/short or quantity contrasts (such as between /ɪ/ and /i/) are important, but quality contrasts are not. L2-influenced vowel qualities are thus acceptable if they follow these principles and are consistent. The NURSE or /ɜː/ vowel is important for intelligibility and should not be substituted.

4) Suprasegmentals: tonic or nuclear stress is important for intelligibility.

All other features, including vowel contrasts based on quality, word stress, weak forms and other suprasegmentals, are ‘non-core’ in that they are not crucial for intelligibility in lingua franca contexts (Jenkins, 2002: 97). In the ELF literature there is occasional ambiguity over what ‘non-core’ means; for example, in summarizing the indications of the LFC, Jenkins, Cogo and Dewey (2011: 287) state that weak forms ‘did not appear to contribute to intelligibility in these interactions and may even have detracted from it’. They are therefore suggesting that the presence of weak forms may have reduced intelligibility. But while this is intuitively plausible, if this were the case then the absence of weak forms – i.e., the presence of full vowels in function words – would need to be included in the LFC, which it is not.

I do not wish to review the LFC findings in detail here, as this has been done elsewhere (see, e.g., Dziubalska-Kolaczyk & Przedlacka, 2005). Word stress is one
‘non-core’ area that subsequent research suggests may have considerable importance (see, e.g., Walker, 2010). Levis (2016: 3) sounds an overall note of caution: although the straightforward recommendations of the LFC have created an impression of ‘certainty’ for some teachers it remains ‘anything but certain for researchers’, and replicatory studies are needed. Deterding (2013) involved a larger data set and a frequency-based ranking of intelligibility-reducing features. As mentioned above, Deterding (2013: 91) concludes that his findings are broadly consistent with the LFC proposals. Figure 1 is based on the findings reported in Deterding (2013) and shows the distribution of intelligibility-reducing features across four categories (consonants, vowels, suprasegmental and others).

![Figure 1](image-url)

**Figure 1.** The number of tokens of pronunciation-related misunderstandings and their allocation to categories (data from Deterding, 2013)

Bearing this overall distribution in mind, evidence of alignment between these findings and those of the LFC comes from Deterding’s (2013: 89–91) summary. Within the consonantal category (92 tokens), the two largest subcategories were the substitution of approximants or nasals such as [l], [n], [r], [w] and [v] (28 tokens) and the loss of the second consonant from initial consonant clusters (20 tokens). This aligns with the LFC’s stipulations regarding the importance of consonants and of initial consonant clusters. There are few areas of contradiction; one is that L-vocalisation (or omission) is seen as unproblematic by the LFC but caused occasional problems in Deterding’s data (4 tokens, including the perception of *also* as *owns*).

Comparing the findings on vowels, Deterding (2013: 91) begins by noting that the quality-related findings of his study are consistent with those of the LFC. Deterding also found, for example, that substitutions of the NURSE vowel could
be problematic (2 tokens out of a total of 15). An additional finding was that substitutions of the TRAP vowel caused occasional problems (3 tokens). Two of these appeared to involve quality (e.g., matter was perceived as middle) while the other involved quantity (pattern was perceived as part or). Other quantity substitutions were occasionally implicated as the cause of intelligibility problems (3 tokens, as when weed was misheard as with).

Deterding’s overall conclusion, however, is that the absence of a distinction between long and short vowels was not a frequent cause of intelligibility problems (2013: 91), suggesting a possible divergence from the LFC findings. It is noteworthy that many ‘new’ varieties of English are characterized by the absence of quantity contrasts between certain vowels (see, e.g., Kirkpatrick, 2007 on African Englishes; Hung, 2000 on Hong Kong English and other Asian varieties). If the ‘failure’ to distinguish these vowels was quite prevalent, and if there were few noticeable problems, then there are some grounds for concluding that the maintenance of certain length contrasts is not a priority in terms of facilitating international intelligibility. But the complexity of the ‘vowel’ category illustrates the problems of presence versus absence, as noted earlier, however, and points to a need for further research and theorization.

In the suprasegmental category, the findings of the two studies are again largely consistent. As with vowels, there are relatively few problems attributable to only this category. The two most frequently-noted subcategories in Deterding’s data were syllabic modifications in the form of missing syllables (4 tokens) and added syllables (3 tokens), and word stress deviations (4 tokens). The word stress data appear to contradict the LFC, but the low number of tokens leaves the status of word stress uncertain. If we assume, as seems likely, that there was a considerable amount of variation in the use or non-use of suprasegmental features, it is possible to conclude that suprasegmentals did not play a major role in the achievement of intelligibility in these interactions – noting, once again, that the absence of problems does not necessarily mean that a feature or category is unimportant.

3. The concept of functional load

In this section I will review the concept of functional load in order to apply it to the findings of lingua franca intelligibility studies. To clarify its meaning I will identify three important distinctions: between static and dynamic senses of functional load, between narrow and broad views of functional load, and between orientations towards presence or absence in the description and explication of features. Some attempt at clarification is needed, because although the concept of functional load is intuitively appealing it is difficult to define precisely (Meyerstein, 1970, cited in
Surendran & Niyogi, 2003: 2). The linguistic version of the concept first emerged in the discussions of the Prague School in the late 1920s, and the *Travaux de Cercle Linguistique du Prague* of 1931 defines functional load as ‘the degree of utilization of a phonological opposition for the differentiation of the diverse meanings of words in a given language’ (cited in Catford, 1988: 17).

The static and dynamic senses of functional load emerge when we consider what is meant by ‘a given language’. In computing the degree of utilization of different phonological oppositions, do we focus on word formation and the lexicon, or on how often these words are actually used? This question arose soon after the introduction of the concept; Catford (1988) notes that Jakobson (1931) was the first to recognize the importance of frequency of occurrence in approaching functional load. Gilner and Morales (2010) point out that features which are used extensively in word formation are not necessarily prominent in language use, and vice versa. For example, the /ð/ segment plays a minor role in word formation as there are few words that contain it, but many of these words (such as *the*, *this*, *that* and so on) occur very frequently. The static/dynamic distinction is pertinent to studies of lingua franca communication, as both the types and frequencies of words may differ from descriptions of the standard language or from accounts of language use among native speakers. It is not necessarily true that the dynamic sense is preferable to the static sense, however, and Gilner and Morales (2010: 136) emphasise their complementarity: ‘each is able to offer information that the other one cannot’.

Perhaps the most common approach to functional load has involved focusing on phonemic contrasts and the work they perform in distinguishing words (see, e.g., Catford, 1987, 1988; Brown, 1988). From this perspective, the more minimal pairs a phonemic contrast serves to distinguish, the greater the possibility for misunderstanding caused by the absence of contrast, and the higher its functional load. The difference between the /ʊ/, /u/ and /ɪ/, /i/ vowel contrasts is a frequently-cited example. The former distinguishes relatively few minimal pairs (as in *full*/fool) and its functional load is relatively low, but the latter has more pairs (e.g., *lid*/lead, *hill*/heel) and a correspondingly higher functional load; the respective values are 0.07 and 0.95 in Catford’s (1988) listings, in which the highest value is 1.00. Computations of this aspect of functional load have been used to formulate pedagogical recommendations (e.g., Brown, 1988; Catford, 1987).

The computations can be refined to include other considerations, such as the frequency with which the members of the pair occur, whether or not they belong to the same word class, and so on (see, e.g., Brown, 1988, 1991; Rischel, 2007). The importance of these considerations is not always agreed upon, but the general robustness and usefulness of the concept was demonstrated by Munro and Derwing (2006). Their study used Catford’s and Brown’s rankings of phonemic contrasts to test a ‘functional load hypothesis’: pronunciation errors involving contrasts with
Functional load revisited

67

a high functional load cause more comprehensibility problems than those with a low functional load. Munro and Derwing reported that there were no significant conflicts in their use of the Brown and Catford lists, and concluded that their study provided preliminary confirmation of the hypothesis.

However, although this approach to functional load – what I shall call the narrow sense – is prevalent in the literature and has useful applications, it is not the only way in which functional load may be conceived. Surendran and Niyogi (2003) point out that the term ‘contrast’ (or ‘opposition’) may be defined more broadly, to encompass non-binary distinctions and classes of linguistic features. Adopting this broad view of functional load and a computational approach, both Gilner and Morales (2010) and Surendran and Niyogi (2003) conclude that ‘consonants do significantly more work in both word formation and language use’ (Gilner & Morales, 2010: 141). This is a finding that can already be applied to retrodictive explanations – i.e., those that seek to explain intelligibility problems in terms of likely causative factors – bearing in mind the caveats noted earlier. The broad sense of functional load is concerned with the general ‘information relevance’ of different features, and reflects the legacy of the information-theoretic approach to communication and functional load developed by Shannon (1951). Shannon was investigating the degree of compression that could be applied to electronic data transmission, but in considering oral communication the question is how much acoustic information can be removed or altered without significantly affecting intelligibility.

This key question leads into the third aspect of functional load discussed here, namely the difference between presence and absence in discussions of the concept (and in general discussions of intelligibility). The distinction has already arisen in the discussion of the findings of lingua franca intelligibility studies in the previous section, and I will examine it in more detail here. One important aspect of the distinction is simply the way that our attention is directed by statements couched in terms of either presence or absence. The statement that consonants ‘do significantly more work’ is based on an assumption of presence, but it can also be usefully restated in terms of absence: removing consonants will generally have a greater effect on intelligibility than removing vowels, for example.

This aspect of the presence/absence distinction may seem obvious, but it is worth pursuing the point for the sake of clarity in discussing findings and phenomena. The presence/absence distinction is highly relevant to considerations of international (or lingua franca) communication. Speakers may come from an environment in which certain features or contrasts are often absent (e.g., London or Nigeria, in the case of H-dropping). Or certain contrasts may be difficult to acquire for certain speakers, because of the effects of the L1 (e.g., /n, l/ for Cantonese speakers). Equally, fluent speakers usually employ vowel reduction instead of
using full vowels. If these speakers enter an environment in which listeners are used to relying on these sounds and contrasts, intelligibility problems may arise—and whether we talk about the causes in terms of presence or absence is important. The absence of nativelike features in L2 speech continues to be a preoccupation of much research and pedagogy, but the seemingly ‘natural’ presence of vowel reduction and other suprasegmental phenomena in native speaker pronunciation may be a problematic absence from the perspective of some non-native listeners.

Another aspect of the presence/absence distinction is that the absence of a contrast from a community repertoire does not mean the contrast has no functional load. For example, Abbott (1991: 223) narrates how he overheard two speakers in East Africa experiencing a momentary intelligibility problem: the listener was uncertain whether the speaker’s intended word was distraction or destruction, and both were contextually possible. Though not explained as such by Abbott, the likely phonetic cause is that the /æ, ʌ/, TRAP/STRUT vowels are often merged in East African varieties of English (Mesthrie, 2010: 522). Because of the loss of information, the listener resorted to an alternative communicative strategy—an exaggerated ‘spelling pronunciation’—in order to establish the speaker’s intended meaning. The absent contrast has a demonstrable functional load, although it may not be the same in all varieties of English (Kirkpatrick, 2007: 18). When ‘information loss’ is significant, speakers can attempt to deal with it in various ways, as this example demonstrates.

3.1 Functional load as a functionalist concept

The theoretical foundations of functional load are further strengthened by being placed within a wider ‘functionalist’ approach to language and communication (see, e.g., Bardovi-Harlig, 2007; Bybee, 2001; Bybee & Hopper, 2001; Givón, 1995). Bardovi-Harlig (2007: 59) claims functional load as a functionalist concept, and offers the following example of its application to grammatical features: if a sentence uses only past-tense verb morphology to mark pastness, the functional load of such verb morphology is high (assuming that ‘pastness’ is functionally relevant in the particular situation). If an adverb such as yesterday is also used in the sentence, the functional load of both devices decreases because the removal of either would still allow ‘pastness’ to be discerned. This interpretation of functional load reflects the ‘broad’ sense of the term discussed above, and involves presence/absence in that it can be applied to both the addition and the removal of features. Similar examples can be given for phonology, including the more segments a word has, the lower the functional load of each segment (assuming that similar positions are compared). Deleting the final consonant from the word statement will cause fewer problems that deleting the final consonant from the word state, for
example. This perhaps explains the finding of Kennedy (2012: 57), who found that when listeners had difficulty understanding multisyllabic words, more than one segment was usually involved.

The functionalist approach stresses ‘the centrality of meaning and function in influencing language structure and language acquisition’ (Bardovi-Harlig, 2007: 57). As mentioned in the previous section, the influence of function is seen as occurring at inter-related levels of language (e.g., lexis, grammar and phonology) and across inter-related temporal frames (e.g., language learning, interaction and language change). The frequency with which features occur is viewed as being of central importance (Bybee, 2001). Language learning is essentially a process of registering frequencies and patterns of co-occurrence, and relating these to possible meanings; Shannon (1951: 54) remarks on the fact that ‘anyone speaking a language possesses, implicitly, an enormous knowledge of the statistics of the language’.

In the temporal frame of interaction, observable patterns of frequency and distribution are not seen as the effects of a pre-existing system but rather reflect ‘the individual speakers’ past experience of these forms, and their assessment of the present context, including especially their interlocutors, whose experiences and assessments may be quite different’ (Hopper, 1998: 156). This seems particularly applicable to lingua franca interactions, in which speakers and listeners may have different repertoires but still need to find sufficient common ground for communication. Speakers bring their past communicative habits to the table, and although their ‘past experience’ may vary greatly there are also many areas of similarity; this is presupposed by the very idea of lingua franca communication. However, if speakers’ past experience leads them to omit or modify features that listeners are used to relying on, then intelligibility problems will occur. This suggests the continuing relevance, for research and pedagogy, of L1 influence – something that ELF [English as a Lingua Franca] research often downplays. Jenkins (2005: 152), for example, contends that ‘much of an ELF speaker’s accent is legitimately free to transfer pronunciation features from the L1’.

In the longer-term temporal frame of language change, Bybee (2001: 2) argues that ‘the frequency with which certain words, phrases or patterns are used…[has] an impact on phonological structure.’ The resultant structural patterns in turn affect the frequency with which features are used, in a dynamic process. But ‘structure’ is not taken to mean fixed, underlying or abstract forms, as in classical structuralist approaches to language. Bybee and Hopper (2001) argue against the concept of ‘pre-existent holistic grammar as the most important pre-requisite for communication'. They quote instead from the work of Roy Harris:
language is envisaged, not as something which exists as a system over and above the communicative situations in which it is manifested, but as a cumulative product of such situations which can be variously exploited to provide a basis for their subsequent renewal (Harris, 1990 [1978]: 150)

This view of ‘structure’ or ‘system’ is taken further by researchers who see language as a complex adaptive system, or CAS: ‘[t]he structures of language emerge from interrelated patterns of experience, social interaction, and cognitive processes’ (Beckner et al., 2009: 2). Seen from this viewpoint, functional load is not an inherent property of sounds or systems; rather, it reflects the way that language users process language in particular situations, based on their past experiences. The term ‘functional load’ is in fact rather misleading, as it encourages us to think of information being ‘carried’ by linguistic features.

Linguistic features do not themselves carry information, however (see Reddy, 1993; Croft, 2000). Although functional load may enable us to make some useful generalisations about the features that are relied upon to achieve intelligibility, and although there may be a certain systematicity in the way that certain features recur, functional load is not a fixed property of features or of a disembodied language system. Discussions of the functional load of features must always take place with reference to the context in which they are being used. With this caveat in mind, in the next section I shall apply the concept of functional load to the findings of lingua franca intelligibility studies.

4. Functional load and the findings of lingua franca intelligibility studies

Researchers in the lingua franca field have so far shown considerable reluctance to consider functional load as something that might help to explain their findings. Deterding (2013) does not mention it at all, and Jenkins (2000: 144) briefly entertains the notion before placing it in the same category as ‘intuition’ and ‘frequency counts’, and dismissing it in favour of ‘solid evidence’. One reason for this may be that the static and narrow senses of functional load appear to emphasize ‘system’, and are seen as inadequate for the task of explaining the fluid and emergent aspects of lingua franca interaction. The apparent reliance on a pre-existing system is also seen as privileging the existing communicative practices of native speakers, rather than allowing for the possibilities of innovation and diversity. But as explained above, acknowledging the ‘dynamic’ and ‘broad’ views of functional load should be able to overcome these objections. Placing the concept within a functionalist approach to language acquisition and use aligns it with emergence, as well as with system. In this section I will consider the question of how far functional load, in
the various senses of the term discussed above, is able to explain the findings of lingua franca intelligibility studies.

As already mentioned, functional load provides a possible explanation for the finding that consonantal modifications – deletions, substitutions and additions – are the major cause of intelligibility problems in lingua franca interactions. Surendran and Niyogi (2003) calculated that the functional load of consonants was greater than that of vowels in English, as well as in Dutch and German. Some insight into the ‘structural’ reasons for this can be gained from the study of Gilner and Morales (2010), who analyzed the 10,000 most frequent words in the British National Corpus (i.e. 10,000 word types, and 9,174,650 word tokens). Firstly, consonants occur more frequently in both types and tokens; the ratios of vowels to consonants were 1:1.78 and 1:1.54 respectively. Secondly, and related to this, the CVC syllable structure is the most common in terms of both types and tokens (36% of the 11,747,726 syllable tokens were CVC syllables). Thirdly, monosyllabic words are the most common, accounting for 79% of tokens.

It has also been frequently observed that while consonants are relatively stable across English varieties and dialects, vowels show greater variability (see, e.g., Rogerson-Revell, 2011: 110). Listeners are therefore more accustomed to dealing with vowel variation. Putting these facts together and taking a broad view of functional load, one would therefore expect consonantal modifications to have a more serious effect on intelligibility. This is particularly the case for modifications in initial position: the beginnings of words play an important role in lexical access (Bent, Bradlow & Smith, 2007; Marslen-Wilson & Zwitserlood, 1989). Initial segments generally have a greater functional load because modifications in this position open up a greater range of possibilities for the listener, compared with modifications occurring later on in the word. As Shannon (1951: 55) puts it, the beginning of words and syllables is where ‘the line of thought has more possibility of branching out’. Syllable initial consonants in strong syllables were found to be ‘particularly important to the listener’ by Zielinski (2008: 81).

The data in Deterding’s study illustrate the operation of these factors. Within the three consonantal classifications adopted by Deterding, the most common environment for intelligibility-reducing modifications was in word-initial position. The most numerous subcategories of consonantal tokens were the substitution of /n/ or /ŋ/ with [l], affecting the words nearby, hanging, noisy, north, next (twice) and northern; the substitution of /l/ with [r] in the words club, black, literature, Labuan, fluently, and like; and H-dropping in higher, has (followed by rise, where there is also a grammatical error), hold, handle, hot cake, and help (us). The prevalence of word-initial modifications as causes of intelligibility problems can easily be gauged from these examples. More specifically, deletion and substitution in word-initial clusters had serious consequences for intelligibility. There were 20
tokens in this category, and the consonants /l/ and /r/ were again frequently implicated, as in the cases of flaming, plough, close, floating, plastic and plant, and afraid, phrase, process, throw, present, break and treaty.

Looking at the particular consonants that were involved in these intelligibility incidents, functional load provides an additional layer of explanation. The consonants [n], [l] and [r] occur relatively frequently, appearing in the upper reaches of the various consonant frequency rankings that have been produced over the years (e.g., Denes, 1963; Gilner and Morales, 2010; Mines et al., 1978). Although these frequency rankings do not take account of the position of occurrence in the syllable, speakers who consistently make modifications to these sounds will therefore make them relatively frequently. Intelligibility problems are therefore more likely to occur, simply because modifications occur more often and affect more words. In fact, in considering the data in Deterding (2013) one must take into account frequency effects, in addition to intelligibility effects. The large number of tokens of misunderstanding involving [n] and [l] reflects the frequent occurrence of contexts for these sounds and certain speakers’ tendencies to modify them, as well as the deleterious effects that modification had on intelligibility.

The narrow sense of functional load (focusing on binary contrasts and minimal pairs) also appears to have some explanatory power. The contrasts between /n, l/ and /ɹ, l/ are generally agreed to have high functional loads (Brown, 1988; Catford, 1988). On the other hand, the data suggest that the narrow or ‘minimal pairs’ approach to functional load may not be particularly relevant, simply because most of the examples of misunderstanding listed by Deterding (2013) do not involve the confusion of minimal pairs. Of the 28 tokens of approximant or nasal replacement, only two appear to involve minimal pairs: reverse heard as we worse (the pair is verse/worse), and want the wall heard as won the war (the pair is wall/war). The other 26 tokens involve either near-minimal pairs, such as when football was heard as food burn, or the perception of non-words (noisy as loisy, next as lext, hanging as haling).

For insight into the types of misunderstanding that occurred, the psycholinguistic study of non-native listeners carried out by Weber and Cutler (2004) is useful. Weber and Cutler believe that a large part of the ‘notorious difficulty’ of listening to speech in a non-native language arises because of the unwanted activation of ‘spurious competitor words’ (p. 3). Moreover, these competitors are not limited to minimal pairs. Partial matches (or near-minimal pairs) are a serious problem because they lead to temporary competitor activation: ‘of kettle when catalog is heard, of belly when balance is heard, of pencil when panda is heard, and so on through the vocabulary’ (Weber & Cutler, 2004: 20). The activation of non-words also occurs. Although such activations may only be temporary, these forms
of competition have been found to cause a measurable delay in word recognition (Norris, McQueen & Cutler, 1995, in Weber & Cutler, 2004: 20).

Looking at the data more generally and applying what has been learned about functional load, it is possible to make some tentative explanations and generalizations. Modifications to consonants, including deletion (as in H-dropping) and substitution (replacing [l] with [n]), create the potential for intelligibility problems because consonants generally have a higher functional load. The chance of these modifications becoming actual problems – their severity – will be greater if they occur in positions that have a high functional load (e.g., word-initial position), or if there are modifications to other nearby sounds. If the target sounds affected by these modifications are frequently-occurring, and if the speaker’s tendency to modify them is consistent, then the frequency of intelligibility problems in the data will also be higher. Both severity and frequency can therefore be argued to have a functional basis.

If functional load is partly context-specific, as was argued above, it is informative to see how the principles apply to other contexts and data sets. One example is briefly considered here: Bond (1999) collected a dataset of almost 900 ‘tokens of misperception’ (or ‘slips of the ear’) experienced by native English listeners. To a certain extent, the overall picture closely resembles that provided by lingua franca studies. Bond found that tokens of misunderstanding were more often related to the production or perception of consonants, rather than of vowels, and that the majority of these errors occurred in initial position (p. 23). Most errors were complicated rather than simple, in that they affected more than one aspect (p. 127).

On the other hand, Bond’s data also reveals some differences between the sources of native and non-native problems. For example, misperceptions caused by consonant cluster modifications were more likely to occur in final position, rather than in initial position. However, there is a fairly obvious frequency-related explanation: the modification of initial clusters is rare among adult native speakers, who have internalized patterns of simplification based on their experience as listeners and speakers. The modification of final clusters is a universal phenomenon, on the other hand (Schreier, 2005). There were more tokens of final cluster modification, and hence a greater proportion of the tokens of misunderstanding were attributable to them, compared with the findings of lingua franca studies. Again, it is important to consider presence and absence, and to take account of frequency effects as well as intelligibility effects when considering this kind of data.

The differences between native speaker and non-native speaker intelligibility problems further illustrate the fact that functional load is not a set of universal properties. There is no fixed core of features that maintains intelligibility for all users, and the sources of intelligibility problems in particular situations are not exactly the same. But neither are they entirely unpredictable. There are some
common principles at work, and the various senses of functional load may assist in characterizing and explaining their importance in any given situation.

5. Conclusions: Functional load in research and pedagogy

This article has combined insights from different fields in order to arrive at a theoretical elaboration of the concept of functional load. Among other things, this established the importance of moving beyond the narrow sense of minimal pairs. The dynamic, broad senses of functional load were linked to wider ‘functionalist’ approaches to language and communication, and to an awareness of frequency effects and emergent structure. Functional load is not a universalizing concept, but it is still capable of suggesting some of the constraints and affordances that shape the contours of communicative practices. While allowing for the fact that intelligibility is an interactive process, the judicious application of functional load may assist in predicting where intelligibility problems are likely to occur, for both native and non-native speakers. Functional considerations can take us a few steps closer to appreciating why ‘different features have different effects on understanding’ (Levis, 2005: 370).

The revisitation of functional load carried out in this article suggests that the concept has some value in approaching the findings of lingua franca intelligibility studies, and there are two main indications. First of all, functional load places the findings of these studies on a firmer theoretical foundation by providing tentative explanations for some of the observed patterns and commonalities, something that has been lacking in previous research. But secondly, considering functional load also tends to subvert the idea that lingua franca communication is substantially different from its implied other, non-lingua franca communication. What the LFC actually shows, from this point of view, is that most of the features and contrasts in existing descriptions are in fact used and relied upon quite widely, and that the vast majority of consonantal contrasts found in so-called ‘standard’ models such as RP and GA are also required for lingua franca communication. It is not therefore clear how the LFC ‘repudiates’ native-speaker norms, as Yazan (2015) claims and as is frequently suggested by the over-polarised rhetoric of some ELF researchers (see MacKenzie, 2014, for commentary).

Finally, this leads into an additional revisitation, that of the pedagogical implications of lingua franca intelligibility studies. As mentioned earlier, the apparent clarity and plausibility of the LFC findings, and their resultant appeal for teachers, have been flagged as a cause for concern (Levis, 2016). While the broad-brush indications of the LFC and associated research – the importance of segmental features, and the importance of consonants – are supported by functional load
considerations, the introduction of the concept also carries several caveats regarding the dangers of overgeneralization and premature application. One of these is the presence/absence distinction: that certain features do not show up as being problematic does not mean they are always unproblematic, or that they do not need to be taught. Considering functional load allows for theoretical evaluations of the possible intelligibility effects of different kinds of modifications, in both segmental and suprasegmental areas.

If the revised concept of functional load is applied to the findings of lingua franca intelligibility studies, the pedagogical implications are fairly clear: assuming that international intelligibility is a priority, certain consonantal modifications that are related to the L1 (such as those affecting the /n, l/ contrast for Cantonese speakers from Hong Kong) or are otherwise associated with particular groups (such as H-dropping for speakers of Nigerian English) should be avoided. There appears to be some scope for functionally-informed, L1-specific or speaker-specific guidelines in pronunciation instruction. Indeed, one indication of Deterding’s data is that the consonantal modifications were not only frequently implicated as the cause of intelligibility problems, but were also disproportionately associated with particular speakers. 17 of the 20 tokens of initial consonant cluster modification came from the Hong Kong speaker, for example, and all 6 tokens of H-dropping came from the Nigerian speaker. To paraphrase Levis (2005: 370), attention to these features, when they occur, may have a disproportionate effect in increasing international intelligibility.

Alternatively, and given the near-intractable difficulties of ‘unlearning’ many of these features, pronunciation instruction could also target the functional adaptations that can be made by listeners. Users of English need awareness of different accents and pronunciation patterns (see Derwing, Rossiter, & Munro, 2002; Lindemann et al., 2016). It is not infeasible to prepare listeners for Nigerian English by making them aware of H-dropping and training them to ‘put sounds back’ during the listening process, for example. Similarly, strategies for listening to Hong Kong English might include trying substitutions of [n] and [l] in cases of momentary difficulty (‘loisy’ = noisy, ‘lexi’ = next). If intelligibility problems do arise – and they have to be seen as inevitable – listeners and speakers need general communicative strategies for dealing with them. These might include asking for repetition, clarification or adjustments to speaking rate, all in a face-saving way. Misunderstandings are not aberrations, as Bailey (2004: 410) points out; they are ‘normal instances of the negotiation of social and linguistic lives’.
Acknowledgements

I thank the Editor, John Levis, and the two anonymous reviewers for their comments on an earlier version of this paper.

References


Bybee, J., & Hopper, P. J. (2001). Introduction to frequency and the emergence of linguistic structure. In J. Bybee & P. J. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 1–24). Amsterdam: John Benjamins. doi: 10.1075/tsl.45.01byb


doi:10.1002/9781444318159.ch25


doi:10.1016/j.system.2006.09.004


doi:10.1075/jslp.1.1.01mun


doi:10.1111/j.1467-971X.2008.00568.x


doi:10.1111/j.1467-1770.1974.tb00502.x


doi:10.1057/9780230513327


**Author’s address**

Andrew Sewell
Department of English
Lingnan University
8 Castle Peak Road
Tuen Mun, NT
Hong Kong
China