The following publication Ogura, M., & Wang, W. S. (2022). Ambiguity resolution and the evolution of homophones in English. In B. Los, C. Cowie, P. Honeybone & G. Trousdale (Eds.), English Historical Linguistics: Change in Structure and Meaning. Papers from the XXth ICEHL (pp. 61-90). Amsterdam: John Benjamins is available at https://dx.doi.org/10.1075/cilt.358.03ogu.

Ambiguity Resolution and the Evolution of Homophones in English

Mieko Ogura[1] & William S-Y. Wang[2]

[1]Linguistics Laboratory, Tsurumi University, Yokohama

[2]Research Centre for Language, Cognition, & Neuroscience, Hong Kong Polytechnic University, Hong Kong

Abstract

Based on a quantitative study of the evolution of homophones in English, we present an argument about why homophones occur. Zipf's law, which states that word frequency decreases as a power law of its rank, can be seen as the outcome of form-meaning associations, adopted in order to comply with listener and speaker needs. This implies that one form can correspond to many meanings (i.e., polysemy and homophony). We argue that homophony is a desirable feature in communication systems, is stable, and increases through time. When a large number of homophones emerge, however, an impetus to avoid homophones comes into play. We suggest that the evolution of diatones is a case of the avoidance of homophony. Related to this, we examine the neural substrates of bisyllabic noun-verb homophones, using near-infrared spectroscopy. We show that noun and verb categories are represented in different neural substrates in the left hemisphere, and relate this to our historical data, explaining why the actuation of diatone-formation was connected with production in frequent homophones in the 16th century, but was connected with perception in infrequent words after the 17th century.

Keywords

linguistic evolution, homophones, ambiguity, Zipf's law, word frequency, diatones, neural substrates

1. Introductory remarks

The existence and abundance of ambiguity in languages has long intrigued linguists. If we view language as a system to encode meanings as signals, it would seem that language is not ideal, because in an ideal code one signal should correspond to exactly one meaning. Each linguistic form should map bijectively to one meaning, so that listeners do not need to expend effort inferring what the speaker intends to convey.

When there are one-to-many correspondences between form and meaning, as in homophony and polysemy, lexical ambiguities arise. Following Piantadosi et al. (2012), we argue that all efficient communication systems are ambiguous, assuming that context is informative about meaning. In the neural substrates, all meanings become active initially and the context leads to the appropriate meaning, or frequency-based dominance and contextual congruity lead to the appropriate meaning (Zempleni et al. 2007).

Anttila (1989, section 9.5) argues that all languages have homophony to different degrees, and that we can never predict with complete confidence when a community or speaker will find it inconvenient enough to remove it. Even when the avoidance or correction of homophony does occur, there is no way of telling by what mechanism of change it will happen.

In this paper, we will offer an argument for the existence of homophones, based on a quantitative study of the evolution homophones in English. We argue that homophony is a desirable feature in communication system and that it likely results from the ubiquitous pressure for efficient communication. However, when a large number of homophones emerge, a force which aims to avoid homophones comes into play. We examine the evolution of diatones, i.e., noun-verb pairs where the stress falls on the first syllable for the noun but the second syllable for the verb, as a case of the avoidance of homophony. In addition to this, we investigate the neural substrates of homophones and diatones in English.

2. The evolution of homophones

2.1. Why do homophones exist?

Normal communication is not much hampered by homophones, because we generally do not process sentences in isolation from contexts. Our quantitative claims could be criticised on the basis that instances of homophonic clash cause ambiguity and must be examined on a case-by-case basis. However, as shown in Table 1, the number of homophones decreases as the number of words in a homophonous set increases, and the percentage of homophones for each number of syllables decreases as the syllable number increases.

Zipf (1949, Chapter 2) argues that the balance of speakers' and listeners' interests is met in a balance between word frequency and the number of words (see below). He further extends his thinking to the balance between the average number of meanings per word and the word rank, showing that frequent words are more ambiguous. Zipf (1949, Chapter 3) also discusses the relation between word frequency and the length of words in terms of the number of syllables, showing that short words are more frequent. We can observe in the number of syllables and the number of words in the homophonous set in Table 1 (together with the number of syllables and word frequency in Table 2 below), that the smaller the number of syllables, i.e., the shorter the words, the higher the word frequency and the larger the number of homophonous sets. Also, the larger the number of words in a homophonous set, the more ambiguous the homophonous set is, and therefore the more ambiguous the shorter words are.

Piantadosi et al. (2012) quantify Zipf's (1949) proposal for trading off between ease of comprehension and ease of production by means of the concepts of 'clarity' and 'ease'. A clear communication system is one in which the intended meaning can be recovered from the signal with high probability. An easy communication system is one in which signals are efficiently produced, communicated, and processed. There are many factors which likely determine ease for human language: for instance, words which are easy to process are likely short, frequent, and phonotactically well-formed. We can observe the trade-off between two communicative pressures (clarity and ease) in the number of syllables and the number of words in the homophonous set in Table 1. The smaller the number of syllables (i.e., the shorter the word) the easier the word is to produce. The larger the number of words in a homophonous set, the smaller the probability with which the intended meaning can be recovered from the signal.

Based on the CELEX lexical database of English (version 2.5, Baayen et al. 1995), we give, in Table 1, the number of homophones in English (classified according to the number of syllables) in the columns, and the number of words in a homophonous set in the rows. The total number of types classified according to number of syllables is also given in parentheses. 11,980 (22.8%) of the 52,447 types are homophones in Present-Day English. 4,743 (70.2%) of the 6,761 one-syllable words and 4,509 (24.3%) of the 18,564 two-syllable words are homophones. In total, they form 77.2% of all homophones.[1] That is, the majority of homophones are short and easy to produce, and they are frequent words. The 6,761 one syllable words and 18,564 two syllable words make up 48.3% of all 52,447 words. The distribution of homophones suggests a threshold of homophones that can be tolerated: around 20% of all words, which is the boundary of high frequency words (see below and section 2.2).[2]

	1 syl.	2 syls.	3 syls.	4 syls.	5 syls.	6 syls.	7 syls.	8-12 syls.	total
2 words	3068	3888	1792	588	132	8	6	0	9482
3 words	900	477	123	60	3	0	0	0	1563
4 words	460	104	16	0	0	0	0	0	580
5 words	175	40	0	0	0	0	0	0	215
6 words	96	0	0	0	0	0	0	0	96
7 words	28	0	0	0	0	0	0	0	28
8 words	16	0	0	0	0	0	0	0	16
total	4743	4509	1931	648	135	8	6	0	11980
	(6761)	(18564)	(15195)	(7970)	(3000)	(711)	(188)	(58)	(52447)

Table 1: The number of homophones in Present-Day English

Why do homophones occur, despite the fact that, in an ideal code, one signal corresponds to one meaning? Zipf (1949, Chapter 2) suggests this is due to the simultaneous minimization of the two opposing forces from listener and speaker for form and meaning association. One form for all meanings is an ideal code for the speaker, while one-to-one correspondence between form and meaning is an ideal code for the listener. Zipf's law, which states that word frequencies decrease as a power law of its rank, may be seen as the outcome of form-meaning association adopted in order to comply with listener and speaker needs. Arranging signals according to Zipf's law is the optimal solution for maximizing the referential power while respecting the constrain regarding effort for a speaker. Zipf's law implies that there should be one form to many meanings, i.e., polysemy and homophony. Polysemy and homophony are therefore the necessary conditions for symbolic systems (Ferrer i Cancho & Solé 2003).

Zipf argues that the balance of speakers' and listeners' interests is observed in a balance between word frequency and the number of words. If word frequencies are tabulated and the words ranked from most frequent to least frequent, a simple formula describes the relation. This relation is known as Zipf's law. If p_r is the probability of the rth most frequent word, then:

$$Pr = \frac{0.1}{r} .$$

When plotted on log-log coordinates, this function gives a straight line with a slope of -1 (Zipf 1949, Chapter 2, Miller 1981: 107). It shows a heavy-tailed distribution, and we may draw a boundary between the most frequent 2,000 words and the remaining 8,000 words in a set of 10,000 words (see 2.2 below).

Ferrer i Cancho & Solé (2003) provide a formal backing for Zipf's intuitive explanation, showing that the power law distribution arises when information-theoretic difficulty for speakers and listeners is appropriately balanced. Piantadosi et al. (2012) argue that ambiguity can be understood as the trade-off between two communicative pressures which are inherent to any communicative system: clarity and ease (see above). For further discussion on the trade-off between speaker and listener needs in the neural substrates, see section 4.1.

2.2. Word frequencies and homophones in Present-Day English

Table 2 shows the average word frequencies of homophones, and of all words, classified according to their number of syllables, in the CELEX database of English, where the frequency information is taken from the 17.9 million token COBUILD/Birmingham Corpus. We also give the percentage of homophones (for the number of homophones and of all words, see Table 1). Table 3 gives the average word frequencies ranked from the most frequent to the least frequent in the CELEX database, and Figure 1 plots the power-law distribution of word frequencies as a function of word rank on log-log coordinates. This graph shows a similar slope to Zipf's law, which gives a straight line with a slope of -1. We find that within each number-of-syllables group, average word frequency of homophones is higher than all words except the 6 and 7 syllable words. The average frequencies of homophones in 1, 2 and 3 syllable words are in the rank of 1001-2000, 4001-5000 and 5001-10000, respectively. Homophones with 1, 2 or 3 syllables form the majority of homophones (4743+4509 +1931 out of 11980, i.e., 93%), and the average frequencies are in the top 10,000 most frequent words out of 52,447 words. If we draw a boundary of highly frequent words around the top 20% most frequent words (see section 2.1 above), the threshold of homophones could also be expected to be at around 20% of all words, because the homophones tend to be formed with highly frequent words.

There are subtypes of iconicity, isomorphism (that is, the one-form-one-meaning condition) and automorphism (which holds that linguistic elements which are alike semantically should also resemble one another formally; Haiman 1985: 4). The former type of iconicity underlies homophony, while the latter type underlies polysemy. Because of automorphism of iconicity, it is plausible that there are many polysemous words. In WordNet we find 5,196 (46.0%) out of 11306 verbs, and 15,279 (13.3%) out of 114,513 nouns are polysemous. We also find that, the greater the number of meanings, the higher is word frequency within each lexical category (Ogura & Wang 2012a). As for homophony, people avoid homophones because of isomorphism of iconicity. Thus we assume that there is some threshold of homophony. The distribution of homophones in Present-Day English in Table 1 shows this situation.

	1 syl.	2 syls.	3 syls.	4 syls.	5 syls.	6 syls.	7 syls.	8-12 syls.
homophones	2082.5	320.6	101.26	69.91	60.17	1.75	3.5	
all words	2009.8	182.13	72.39	49.86	34.32	13.49	6.63	0.79
% of homophones	70.2%	24.3%	12.7%	8.1%	4.5%	1.1%	3.2%	0%

Table 2: Average word frequencies of homophones and of all words, and the percentage of homophones

Word rank	Average frequency
1-500	26768.54
501-1000	2627.33
1001-2000	1262.48
2001-3000	674.13
3001-4000	430.35
4001-5000	294.31
5001-10000	140.55
10001-15000	53.26
15001-20000	26.24
20001-25000	13.57
25001-30000	6.77
30001-35000	2.97
35001-38731	1.13
38732-52447	0.00

Table 3: Average word frequencies ranked from the most frequent to the least frequent in the CELEX database

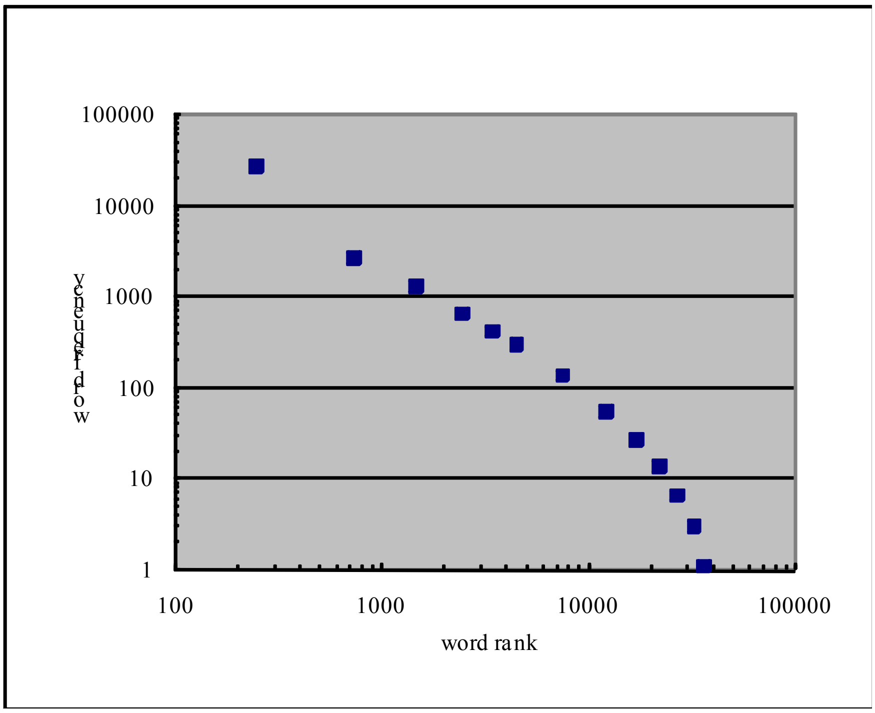


Figure 1: Power-law distribution of word frequencies in the CELEX database

We would like to add that most of the homophones are composed of words from different lexical categories in English. Based on the first 4,919 samples out of 11,980 homophones from the CELEX database, we find that the number of occurrence of words whose lexical categories occur once in each homophonous set is 4,263 (86.7%) of 4,919 homophones. Noun-verb ambiguous pairs form a large portion of these inter-category homophones in English.[3]

We argue that homophony is a desirable feature in a communication system and that it likely results from a ubiquitous pressure for efficient communication. However, when a large number of homophones emerge, the balance of speakers' and listeners' interests observed in the balance between word frequency and the number of words is lost, and the force to avoid homophenes comes into play. One such case is the evolution of diatones in English. Before we discuss the evolution of homophones and diatones after the 16th century (in section 3), we consider the evolution of homophones in Old English (OE) and Middle English (ME).

2.3. Homophones in Old English and Middle English

2.3.1. The stability of OE homophones

The Brooklyn-Geneva-Amsterdam-Helsinki Parsed Corpus of Old English (2000) contains 18,629 types, and 1205 types (6.5%) of them are homophones. The scoring criterion used here, and definition of homophony, is different from that used for the CELEX database of English and the LDC Japanese Lexicon, where a lexeme (a set of forms taken by a single root word) is a basic unit of scoring. However in OE, any differences in forms are counted as different types except scribal variations, because the inflectional system of OE was quite complex and distinct in form in nouns, personal pronouns, adjectives, and verbs (Moore 1969: section 26). Thus, for example, we see *sum* (which has become 'some' in Present-Day English), as having the following 8 homophonous sets: *sum, suman, sume, sumere, sumne, sumon, sumre, sumum* as adjective and pronoun. These are treated as one homophonous set ('some') as an adjective and pronoun in the CELEX database. If a root is used as a representative form, OE has a very restricted number of root words that can stand alone as a word. On the other hand, Present-Day English has little inflection and a tendency to have words that are identical to their roots.

In OE double consonants were typically geminates. In late OE, because of the open syllable lengthening of disyllabic words, long/short consonants and long/short vowels occur in complementary distribution: long consonants after short vowels, and short consonants after long vowels. Thus long consonants are no longer phonemically long, but we use the phonetically long pronunciation, because our calculation for homophonous forms is based on phonetically different forms. Thus, for example, the homophones *al* as adjective and pronoun are treated as different from the homophones *all* as adjective and pronoun (both of which remain as 'all' in Present-Day English).

Zipf (1949, Chapters 2, 3) argues that all words which differ phonetically in the fully inflected form in which they occur should be counted as separate words (thus the forms, *give, gives, gave, given, giving, giver, gift* represent seven different words and not one word in seven different forms). A word is defined as any sequence of successive letters bounded by spaces in the text. He employs the criterion of inflection when he analyzes the rank-frequency distribution of words based on James Joyce's *Ulysses,* and especially in the treatment of highly inflected languages, such as Nootka, German, Gothic and Ælfric's Old English which all use inflections virtually throughout their entire structure. However, he uses the *Thorndike-Century Dictionary*, where lexical units, i.e., words in non-inflected form are listed, when he analyzes the meaning-frequency distribution of words. He states that because of the low degree of inflection of words in Present-Day English, the difference between a rank-frequency distribution of lexical units and that of words in fully inflected form is not considerable.

780 (65%) of 1,205 OE homophones are still used in Present-Day English, and 425 have become obsolete. Table 4 gives the number of types that became obsolete in a set of periods. The dates of obsolescence were checked in the *Oxford English Dictionary, Version 2.0 on CD-ROM (OED2)*. Many of the 144 homophones that became obsolete in the 19th century are archaic, literary or dialectal in Present-Day English. We also find that in 487 (85%) of 572 homophonous sets, at least one type exists in Present-Day English. Thus we may state that OE homophones have been stable over time.

OE	37
12th c.	13
13th c.	68
14th c.	43
15th c.	49
16th c.	23
17th c.	35
18th c.	13
19th c.	144

Table 4: Number of OE homophones that became obsolete

2.3.2. Stability of ME homophones

We have attempted to confirm the above observation about OE homophones in terms of ME homophones, too, based on the *Penn-Helsinki Parsed Corpus of Middle English* (2000). This corpus contains 48,725 types (2,003 clitics are excluded), and we find that 4,653 (9.6%) of them are homophones.

2,691 types (= 1,966 types of OE origin + 725 types of OE origin), which is 57.8% of 4,653 ME homophones (= 2,981 types of OE origin + 1,672 types of OE origin) are used in Present-Day English. Table 5 gives the number of types that became obsolete from the 12th to the 19th century, classified according to OE origin and ME origin. The total of the types of OE origin that became obsolete in the 19th century and those that are still used in Present-Day English are 2,669 types (703 types + 1966 types), which is 89.5% of the 2,981 homophones of OE origin. The total of the types of ME origin that became obsolete in the 19th century and those that are still used in Present-Day English are 1,377 types (652 types + 725 types), which is 82.4% of 1,672 homophones of ME origin. It is not clear how many homophones of OE origin became obsolete in OE from Table 5, but we can confirm that homophones of both OE and ME origin are stable, and that homophones of ME origin add to those of OE origin.

	OE origin	ME origin
12th c.	8	7
13th c.	61	29
14th c.	48	33
15th c.	73	49
16th c.	57	48
17th c.	52	92
18th c.	13	37
19th c.	703	652

Table 5: Number of ME homophones that became obsolete

2.3.3. Word frequency of OE and ME homophones

Tables 6 and 7 give the average word frequencies (ranked from the most frequent to the least frequent) in the OE and ME databases, respectively. Figure 2 plots the power-law distribution of word frequencies in the OE and ME database on log-log coordinates. The slopes of OE and ME graphs are parallel, and they are approximately parallel to the graph for Present-Day English in Figure 1, except the heavy tailed part at the bottom. The average frequencies of OE and ME homophones are 30.7 and 87.9 respectively, and they are high frequency words in the word rank of 251-500 of the OE database in Table 6 and the word rank of 501-1000 of ME database in Table 7, respectively.

word rank	average frequency
1-250	216.53
251-500	29.09
501-1000	15.39
1001-1500	9.29
1501-2000	6.62
2001-3000	4.61
3001-4000	3.24
4001-5000	2.45
5001-10000	1.43
10001-18629	1.00

Table 6: Average word frequencies ranked from the most frequent to the least frequent in the OE database

word rank	word frequency
1-250	1520.26
251-500	164.99
501-1000	78.76
1001-1500	43.21
1501-2000	29.52
2001-3000	19.80
3001-4000	13.30
4001-5000	9.88
5001-10000	5.62
10001-15000	2.89
15001-20000	2.00
20001-25000	1.17
25001-48725	1.00

Table 7: Average word frequencies ranked from the most frequent to the least frequent in the ME database

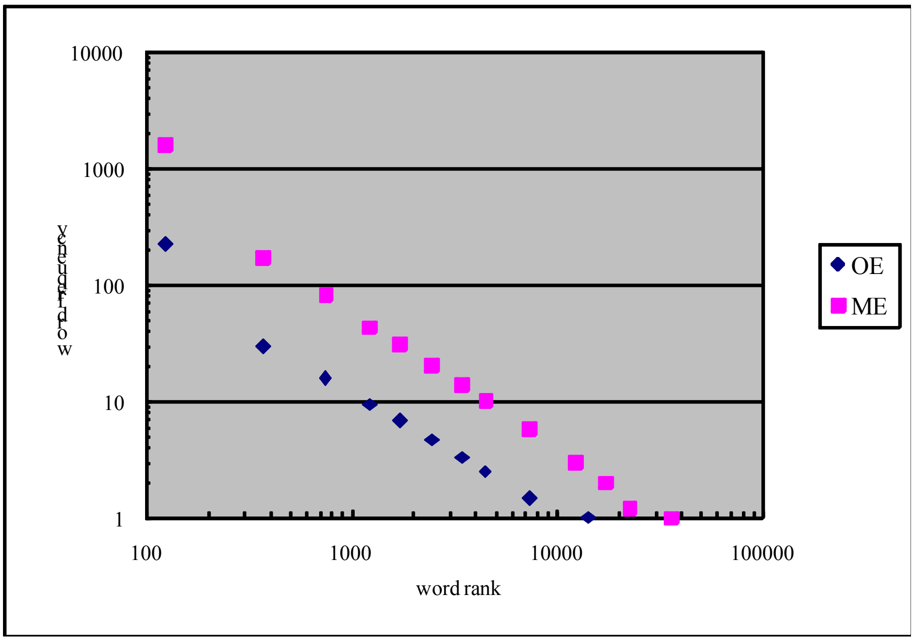


Figure 2: Power-law distribution of word frequencies in the OE and ME database

Table 8 gives the average frequencies of the obsolete words from OE to the 19th century in the OE database. We also give the average frequency of the words that remain in the 20th century. Table 9 gives the average frequencies for the obsolete words from the 12th to the 19th century in the ME database, classified according to OE origin and ME origin. We also give the average frequency for the words that remain in the 20th century. We find that words that became obsolete in OE in the OE database in Table 8 are low frequency words. This is to be expected: rare things are forgotten most easily. From the ME database in Table 9, we find that the frequency of words of OE origin is higher than that of words of ME origin. This is because the low frequency words became obsolete in OE. We also find that there is a strong correlation between dates of obsolescence and word frequencies. Words that became obsolete in the 19th century and those that remain in the 20th century are by far the most frequent words both in terms of OE and ME origin. The average word frequencies from the 12th to the 18th century and those in the 19th and 20th century for words of OE origin and ME origin are correlated by 0.999 of the Pearson correlation coefficient at the 5 per cent level. We note that the average frequency of words that became obsolete in the 13th century in the OE database is high as compared with that in the ME database. This may be due to the change in which the relative pronouns *þe* and *se*, which were frequently used in OE, were replaced by *þat* in the 13th century.

OE	2.27
12th c.	29.08
13th c.	89.06
14th c.	45.58
15th c.	41.04
16th c.	10.65
17th c.	11.23
18th c.	39.77
19th c.	19.45
20th c.	28.98

Table 8: Average frequencies of obsolete words from OE to the 19th century and those of words that remain in the 20th century from the OE database

	OE origin	ME origin
12th c.	17.63	5.71
13th c.	9.10	3.14
14th c.	5.00	3.79
15th c.	47.84	19.27
16th c.	22.16	3.42
17th c.	21.58	3.40
18th c.	5.92	2.97
19th c.	66.15	33.55
20th c.	160.56	22.63

Table 9 Average frequencies of obsolete words from the 12th to the 19th century and those of words that remain in the 20th century from the ME database

3. Evolution of Diatones

3.1. Why do diatones evolve?

As we have argued in section 2, homophony is a desirable feature in a communication system, it is stable and accumulates over time. But when a large number of homophones emerge, a force to avoid homophones comes into play. We suggest that when a large number of homophones emerged in the 16th century, diatones first emerged in order to avoid the existence of homophones. Based on Sherman's (1975) list of noun-verb oxytonic homophones, we give the number of homophones that emerged from the 16th to the 20th century and the total number of homophones in each century in Table 10. We can see that a large number of homophones emerged in the 16th century.

	16th c.	17th c.	18th c.	19th c.	20th c.
number of homophones that emerged	129	33	19	29	-17
total number of homophones	129	162	181	210	193

Table 10: The number of noun-verb oxytonic homophones that emerged from the 16th and 20th century and the total number of homophones in each century

Diatones are noun-verb pairs where the stress falls on the first syllable (paroxytonic) for the noun but on the second syllable (oxytonic) for the verb, e.g. *address, permit, subject, contract*, etc. Sherman (1975) gives a catalogue of a large number of historical sources of diatones. For each N-V pair, he gives accentual evidence from dictionaries and grammars, ranging in date from 1570 (Levins) to 1798 (Jones). There are two early dictionaries: Levins (1570) and Minsheu (1617). Of these two, only Levins provides reliable evidence for any N-V diatones, and he gives three such words: *outlaw, rebel, record*. Twelve years later, Mulcaster (1582) provides evidence for five additional disyllabic diatones (*convert, desert, incense, present, refuse*). This scarcity of sources changes in the 18th century, which sees the publication of nearly 250 dictionaries. Sherman's consultation of 30 of these shows that diatones emerged in the latter half of the 16th century, although he does not have an explanation for their emergence at this particular time.

We find several prerequisite linguistic environments for the evolution of homophones and diatones in the 16th century. In OE there was a predominance of word-initial stress, but as Campbell (1959: 30) notes, there were noun-verb pairs (not homographs) which exhibited stress alternation, with word-initial stress being shown by the noun and non-initial stress by the verb as in *ˈæfþunca* 'source of offence' vs. *ofˈþyncan* 'to displease', *ˈbīgenga* 'inhabitant' vs. *beˈgān* 'to occupy'. It is plausible that these pairs supplied a model for the productive process of diatone-creation which emerged later in Early Modern English (EModE).

In ME and EModE we find the development of two stress rules that have survived to the present: the Germanic Stress Rule and the Romance Stress Rule. This is especially the case with disyllabic words, as is evident from alliterative and rhyme poetry in LME. Abundant double-stressed words occur whether they are native or of Romance origin, or whether they are unaffixed, suffixed, or complex words as shown in: *aˈleye~ˈaleye, boˈsom~ˈbosum, deˈsire~ˈdezyre, folˈlow~ˈfolow, mesˈsage~ˈmessage, lyˈvynge~ˈlyvyng, outˈlawe~ˈoutlawe, acˈcesse~ˈaccess, biˈleve~ˈbilyue, conˈclude~ˈconclude* (Nakao 1978, Lass 1992).

Another linguistic phenomenon is the gradual leveling and loss of inflections from Late OE through Late ME. As a result of this, nouns and verbs fell into one and the same form in many of their inflected forms, and the morphological process of conversion operated more frequently. Yet another linguistic factor is the great influx of French and Latin vocabulary in ME and EModE. In addition to their effect on the stress system, the loanwords themselves provided a rich source of disyllabic words with morphological structure of predominantly prefix plus base.

The present study is based on the 237 diatone data from the late 16th to the early 20th century: Sherman's 147 diatones,[4] and the additional data for the 19th and the 20th centuries from Hotta's (2013) 90 diatones. Hotta collected the data from Sherman's oxytonic portion of the noun-verb isotone inventory, Fudge (1984: 189-92), Jerspersen (1954: 173-82), and a number of dictionaries in the 19th and 20th centuries.

The Appendix to this article shows the evolution of 72 diatones that arose from a large number of homophones from the late 16th to the late 18th century. The 72 diatonic pairs are arranged according to the earliest period at which explicit lexicographic evidence is found for stress alternation in the second column. 16c, 17b and 18c, etc. mean the late 16th, middle 17th, and late 18th century respectively. The third column shows the distinction between N(oun) and V(erb).

The fourth column of the Appendix shows the etymologies based on *OED2*. The abbreviations are as follows: OF. Old French; F./Fr. French; mod. F. Modern French; AF./AFr. Anglo-French; ONF. Old Northern French; L. Latin; ON. Old Norse; OE. Old English; lOE. late Old English; It. Italian; Sp. Spanish; MDu. Middle Dutch; Com. Teut. Common Teutonic; Com. WGerm./WGer. Common West Germanic; ppl. a. participial adjective; < means 'comes from'; conversions are also shown, such as *relapse* (v), which means *relapse* (n) comes from conversion of *relapse* (v). In the Appendix, we can see a great influx of French and Latin vocabulary in ME and Early Modern English (EModE), which are a rich source of disyllabic words. The fifth column shows the date of the first attestation of each noun and verb based on *OED2*. The sixth column shows the period of the first attestation of each isotonic or homophonous N-V pair from which a diatone emerged. We find that a large number of homophones appeared in the 16th and 17th centuries. This is due to a great influx of French and Latin vocabulary in ME and EModE which occurred with gradual inflectional leveling and loss from LOE through LME.

The seventh column shows the patterns of the accentuation, O(xytonic), P(aroxytonic), and D(iatonic) which Sherman lists based on some 30 dictionaries and grammars, ranging in date from 1570 (Levins) to 1789 (Jones). We may assume that originally some isotonic N-V pairs had the stress on the second syllable (O(xytone)), some on the first syllable (P(aroxytone)), and others both (O and P), and that diatones developed from these three patterns. The three patterns developed from the double-stressed disyllabic words where one and the same word manifests final stress at one time and initial stress at another time in ME and EModE. The eighth column shows the word frequency of each diatone pair, based on the *American Heritage Word Frequency Book* (1971).

3.2. Word frequency and diatones

As we follow the course of the evolution of diatones from the 16th to the 20th century, we can see the gradual diffusion of stress-alternation, reaching more and more noun-verb homographic pairs, thereby creating an ever larger number of diatones. The inventory of diatones expanded to 32 by 1700, 72 by 1800, 153 by 1900, 237 by 2000 (here the homograph is counted as diatonic if any source gives a diatonic pattern, even secondarily). If we plot the steady growth of the number of diatonic N-V homographs as a function of time, the resulting graph is a chronological profile of lexical diffusion, as shown in Figure 3. The change is still in progress at the present. Lass (1980: 75-80) argues against homophonic clash or avoidance of homophony as an explanation for certain (usually sporadic) changes. Our work is based on the quantitative examinations of an adaptation through the amplification of and competition between diatones, paroxytone pairs and oxytone pairs, which leads to growth in complexity in language over time.

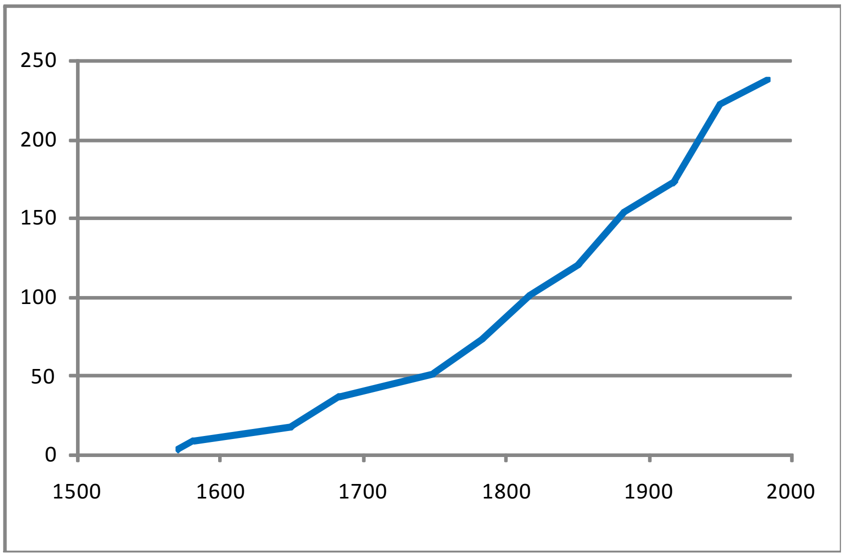


Figure 3: Chronological profile of the lexical diffusion of N-V diatones

Table 11 gives the average word frequencies for the disyllabic diatones and noun-verb homophones from the 16th to the 20th century. The diatone data from the late 16th to the early 20th century are based on Sherman's 147 diatones and the additional data from Hotta's (2012) 90 diatones. The disyllabic noun-verb homophones are based on Sherman's oxytonic noun-verb isotone inventory. Sherman lists 213 pairs (though he claims 215 pairs), from which 3 pairs and 17 pairs changed to diatones in the 19th and 20th centuries respectively, and they are in Hotta's list. The word frequency (frequency-per-million) of each diatone pair and oxytone pair is based on the *American Heritage Word Frequency Book* (1971). The number of occurrences of new diatones and oxytonic homophones are given in parentheses for each century.

	16th c.	17th c.	18th c.	19th c.	20th c.
diatone	44.28(8)	16.99(28)	7.45(36)	4.56(81)	5.27(84)
homophone	13.29(129)	13.14(162)	12.44(181)	11.28(210)	11.21(193)

Table 11: Average word frequencies of disyllabic diatones and noun-verb homophones from the 16th to the 20th century

We find that the formation of diatones started in the high-frequency homophones in the 16th century, but after the 18th century it gradually diffused from the low-frequency homophones. The 17th century is the transitional period. As shown in the Appendix, most of the diatones that appeared in the late 16th (16c) to middle 17th (17b) century have their sources of both noun and verb or either of them in ME and have higher frequency than those that appeared after the late 17th century (17c), whose sources are mostly French and Latin words that were borrowed after the 16th century.

Just after a large number of homophones emerged in the 16th century, the avoidance of homophony strongly operated in the highly-frequent homophones, because, we argue, speakers aimed to avoid homophones intentionally to keep the distinction of the forms and meanings, which were kept distinct in ME before they became homophonous in the 16th century. But in the homophones which remained as they are in the 16th and 17th century, diatones evolved from the less frequent words. Diatone formation is perceptually or cognitively motivated change by the listener, and it is in the less frequent words that the change starts (Phillips 1984).[5]

4. Neural Bases of the Evolution of Homophones and Diatones

4.1. Neural substrates of bisyllabic noun-verb homophones in English

Sereno & Jongman (1995) investigate systematic acoustic differences in bisyllabic noun-verb ambiguous pairs which do not exhibit stress differences. 5 speakers read 16 bisyllabic words in both noun and verb contexts. The 16 bisyllabic homophones consist of four categories: 4 pairs which occur more frequently as forestressed nouns (*favor, poison, practice, struggle*); 4 pairs which occur more frequently as backstressed nouns (*control, debate, dispute, report*); 4 pairs which occur more frequently as forestressed verbs (*handle, notice, rescue, welcome*); and 4 pairs which occur more frequently as backstressed verbs (*embrace, escape, neglect, reply*).

Sereno & Jongman (1995) find that stimuli that are more frequent as nouns in English (e.g. *poison, debate*) show significantly different amplitude ratios than word stimuli that are more frequent as verbs (e.g. *notice, escape*) (pp. 68-69). In other words, the amplitude of the first syllable of word forms that are usually used as nouns is higher than that of words usually used as verbs, regardless of whether the form in question is being pronounced as a noun or a verb (Phillips 2014: 89). Highly-frequent nouns produced a large number of isotonic paroxytone pairs due to speakers' ease of producing those words. Sereno & Jongman (1995) conclude that the significant dominance effects suggest that speakers maximize the difference between noun- and verb-dominant words in conformity with the lexical distribution of English in which the majority of bisyllabic nouns are stressed on the first syllable and the majority of bisyllabic verbs on the second syllable (p.69). The above noun-verb homophonous pairs except the pair for *struggle* emerged in the 16th century, and we may assume that 16th century English shows a similar situation in terms of homophony to Present-Day English. The homophonous pair *struggle* appeared in the 17th century.

We investigate the neural substrates of bisyllabic noun-verb homophones in English using near-infrared spectroscopy (NIRS), inquiring into their interaction with speech production based on the acoustic data by Sereno & Jongman (1995). We used 16 bisyllabic noun-verb ambiguous pairs from Sereno & Jongman (1995). The experimental sentences were constructed such that the noun-context and the verb-context conditions were minimally different prior to the critical ambiguous word, by including sentence initial phrases like *I had/wanted/needed/planned/continued a (no, the)/to* (Fedorenko et al. 2012). Note that the grammatical categories of the words are never ambiguous in the sentence contexts. They are similar in meaning.

The stimuli were presented randomly to six adult (both male and female) native English speakers. The subjects were instructed to listen to the auditorily presented sentences silently. All the sentences were within 5s in length. The interval between the sentences was 19s to allow the hemodynamic response to return to the baseline before initiating the following trial.

The changes in hemoglobin (Hb) concentrations and their oxygeneration levels in the frontal and temporal lobes in the left hemisphere and the right homologous areas were recorded using NIRS systems (ETG-4000; Hitachi, Tokyo, Japan). NIRS measures the Hb concentration changes of the optical paths in the brain between the nearest pairs of incident and detection probes (Watanabe et al. 1996, Yamashita et al. 1996). The system emits two wave-lengths, approximately 780 and 830 nm, of continuous near-infrared lasers, modulated at different frequencies depending on the channels and the wavelengths, and detected with the sharp frequency filters of lock-in amplifiers (Watanabe et al. 1996). The probe geometry is represented in Figure 4. The probes are placed in their respective areas with an emitter-detector separation length of 25 mm (Watanabe et al. 1996, Yamashita et al. 1996). This separation enables us to measure hemodynamic changes in the brain 2.5-3 cm deep from the head surface, which corresponds to the gray matter on the outer surface of the brain (Fukui et al. 2003).

The probe pads were positioned onto the subjects' bilateral frontal and temporal areas as shown in Figure 4. Green squares (1-22 without circles round) indicate the channel positions. Red probes (probes 11-18, 21-28 in grey circles) are emitters, and blue ones (probes 11-18, 21-28 in black circles) are detectors. The numbers of channels 1-22 correspond to the numbers 1-22 of the time course of hemoglobin responses in the left and right hemispheres in Figures 5-8.

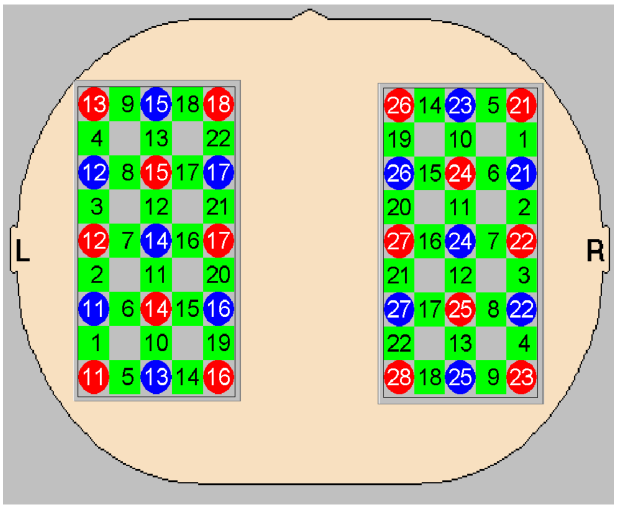


Figure 4: Probe arrangements

Figure 5 shows the average values of the oxy-Hb changes of the 6 English subjects in the frontal and temporal lobes in the left hemisphere and the homologous areas in the right hemisphere for the forestressed nouns in red (solid) lines and the forestressed verbs in blue (broken) lines in noun-dominant words in the range of ±0.10 mM*mm. We find anatomicalbehavioral correlations, with a left frontal cortical network activated for verbs, which is marked with a solid circle, and greater activation in the temporal regions for nouns, which is marked with a broken circle. In the neuroimaging literature, there is a substantial corpus of studies asking whether nouns and verbs are represented in different neural substrates, or whether both categories are processed in the same anatomical area but with functional differences between them (Crepaldi et al. 2011, Tsigka, et al. 2014, Vonk et al. 2015).

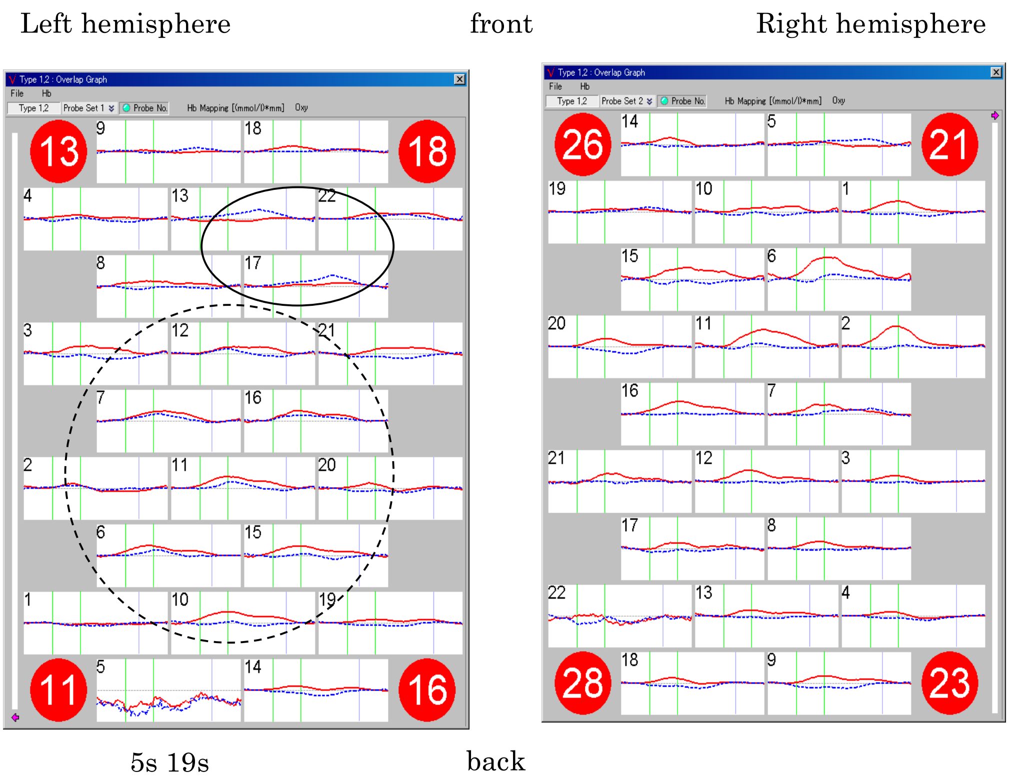


Figure 5: Noun-dominant forestressed nouns and verbs

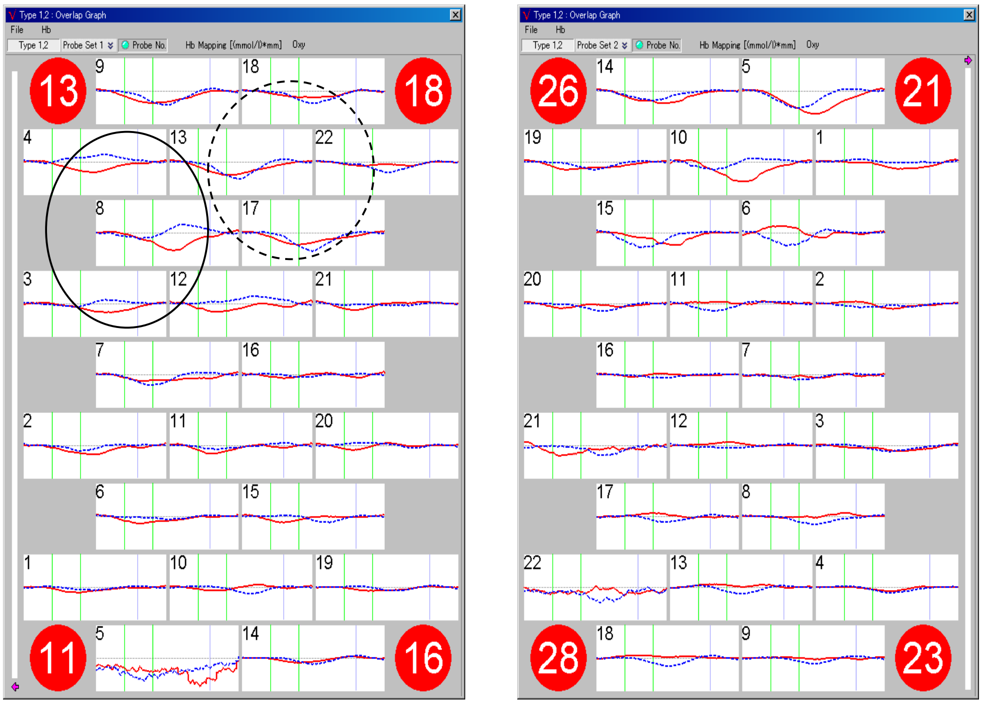
Our results show that noun and verb categories are represented in different neural substrates in the left hemisphere. We argue, following Shapiro & Caramazza (2003), that information about a word's grammatical category is represented independently of its meaning. Semantic information about nouns and verbs is stored in the frontal and temporal lobes, in which categorical distinctions are explicitly represented when nouns and verbs are activated: there is greater activation of nouns than verbs in the temporal regions and greater activation of verbs than nouns in the frontal regions.

Statistically significant differences are observed between nouns and verbs by the t-test at the $p \leq 0.05$ level. In the right hemisphere, the activation of highly frequent nouns in noun-dominant pairs occurs in the frontal and temporal regions. We note that highly frequent nouns activate in the frontal lobe and temporal lobe in the right hemisphere, but less frequent verbs that show greater activation than the highly frequent nouns in the frontal lobe in the left hemisphere do not activate in the frontal lobe in the right hemisphere. We may assume that when highly frequent words activate in the left hemisphere, the activation occurs in the homologous area in the right hemisphere.

Figure 6 shows the verb dominant backstressed nouns in red (solid) lines and verbs in blue (broken) lines. Verb-dominant forestressed nouns and verbs show similar patterns of activation. We find the activation of both forestressed and backstressed verbs in the frontal lobe of the left hemisphere, which is marked with a solid circle. Furthermore, we find the decrease to the negative value in the oxy-Hb for both forestressed and backstressed verbs in the frontal lobe, which is marked with a broken circle. Task-induced deactivation occurs because certain types of neural processes active during passive states are interrupted when subjects are engaged in effortful tasks (Binder et al. 2009). We note that the deactivation occurs in the area where the activation of less frequent verbs in the forestressed noun-dominant pairs occurs (see Figure 5). In the right hemisphere the activation of highly frequent verbs in verb-dominant pairs occurs both in positive and negative values in the frontal regions. We find statistically greater activation of verbs than nouns by a t-test at the $p \leq 0.05$ level in stressed syllables in the frontal lobes in verb-dominant pairs.

We argue that listeners try to maximize the difference between noun- and verb-dominance for the ease of perception, though the amplitudes in stressed syllables in verb-dominant pairs are lower than the amplitude of the first syllable of the forestressed noun-dominant pairs in production. We suggest that the cortical representation of speech does not merely reflect the external acoustic environment, but instead gives rise to the perceptual aspects relevant for the listener's intended goal. In 16th century English, we assume that the noun-verb pairs which exhibit stress alternation from OE and ME (see section 3.1) supplied a model for maximizing the difference between noun- and verb-dominant pairs. Here we can see the trade-off between the speaker's ease of production and the listener's ease of perception of homophones.

Left hemisphere front Right hemisphere



back

Figure 6: Verb-dominant backstressed nouns and verbs

Ogura & Wang (2018) have further suggested that nouns and verbs are originally processed in the same anatomical area of the brain when the basic word order of a language is SOV, and discourse organization is closely interwoven with syntactic organization as shown in OE and Japanese. When the SOV order is changed to SVO through embedding, a strictly syntactic organization of the clause appears, and the noun/verb distinction is represented in different neural substrates in the brain, as shown after ME (for details, see Ogura & Wang 2012b, 2014, 2018).

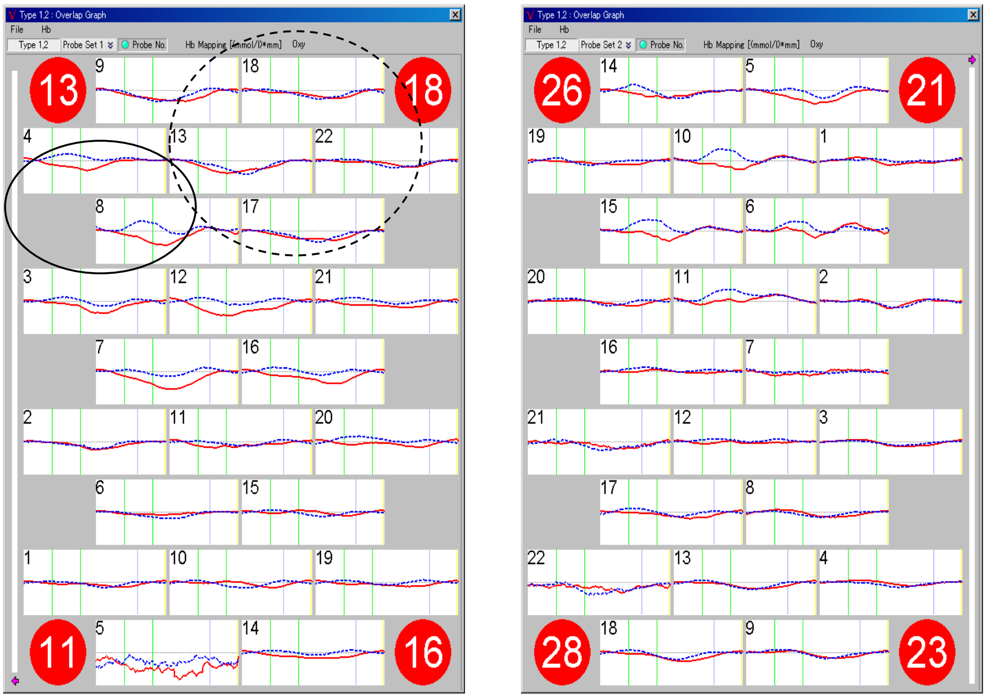
4.2. Neural substrates of diatones

In the experiment, the sentence pairs are organized in terms of noun-plus-verb frequency: four high-frequency pairs (*a subject, to subject; a project,* *to project; a present, to present; progress, to progress*), and four low-frequency pairs (*a compress, to compress; an imprint, to imprint; a digest, to digest; an insult, to insult*). Phillips (2014) demonstrates noun-plus-verb frequency is more influential in the shift to diatones than noun or verb frequency alone.

Figure 7 shows the average values of the oxy-Hb changes for the six English subjects in the frontal and temporal lobes at the 22 channels in the left hemisphere and the homologous areas in the right hemisphere for high-frequency diatones: nouns are shown with red (solid) lines and verbs with blue (broken) lines. Figure 7 shows the activation of verbs in the frontal regions in the left hemisphere, both with positive and negative values, which is circled in a solid and broken line respectively. In the right hemisphere we find the activation of verbs in the frontal regions.

Figure 8 shows the average values of the oxy-Hb changes for the six English subjects in the frontal and temporal lobes at the 22 channels in the left hemisphere and the homologous areas in the right hemisphere for low-frequency diatones: again, nouns are shown with red (solid) lines and verbs with blue (broken) lines. Figure 8 shows the activation of verbs in the frontal and temporal areas in the left hemisphere, where semantic information about nouns and verbs are stored. This is marked with a solid circle. In the right hemisphere there is little activation of verbs. It is interesting that the activation of verbs occurs categorically in the high-frequency diatones, while it occurs semantically in the low-frequency diatones in the left hemisphere.

Left hemisphere front Right hemisphere



back

Figure 7: High-frequency diatones

Left hemisphere front Right hemisphere

back

Figure 8: Low-frequency diatones

Categorical activation of the verbs of diatones in the frontal lobe reflects speakers' intention to keep a distinction between forms and meanings in the production of high-frequency homophones, while the semantic activation of the verbs of diatones in the frontal and temporal lobes reflects perceptually or cognitively motivated change by listeners for low-frequency homophones. The neural substrates of the diatones explain why the actuation of diatone-formation was connected with production in highly frequent homophones in the 16th century, but was connected with perception in infrequent words after the 17th century.

5. Concluding Remarks

Based on a quantitative study of the evolution of homophones in English, we have shown why homophones occur. Zipf's law, which states that word frequency decreases as a power law of its rank, can be seen as the outcome of form-meaning associations adopted in order to comply with listeners' and speakers' needs. This implies one form to many meanings, i.e., polysemy and homophony. Homophony is a desirable feature in a communication system, and it is stable and increases through time. But when a large number of homophones emerge, a force to avoid homophones comes into play. We have suggested that the evolution of diatones is a case of avoidance of homophony.

Furthermore, we have examined the neural substrates of comprehension of bisyllabic noun-verb homophones in English. The evolution of homophones is a result of the interaction between a speaker's production and a listener's perception, and the cortical representation of speech does not merely reflect the external acoustic environment. We have suggested that noun and verb categories are represented in different neural substrates in Present-Day English. We have further shown that the activation of verbs occurs categorically in high-frequency diatones, while it occurs semantically in low-frequency diatones in the left hemisphere. The neural substrates of the diatones explain why the actuation of diatone-formation was connected with production in frequent homophones in the 16th century, but was connected with perception in infrequent words after the 17th century.

Notes

1. We have also analyzed the LDC Japanese Lexicon (1997) and found similar results in Present-Day Japanese: 8,827 (17.2%) of the 51,274 types are homophones; 145 (58.3%) of the 252 one-syllable words and 1.232 (40.9%) of the 2,946 two-syllable words are homophones (Ogura & Wang 2006).

2. Strang (1980) collected some 1700 monosyllabic words from the *OED*, and found that the level of exploitation of homophones is higher than might be thought tolerable in several forms: /biː/, /bɛː/, /bei/, /bʌt/, /bʌk/. For example, /bʌt/ has had 21 homophones, including 16 nouns and 4 verbs, with 12 surviving till the 19th century.

3. In Japanese, however, most of the homophones are composed of words from the same lexical categories. The number of occurrences of words whose lexical categories occur once in each homophonous set is 365 (4.1%) of 8,902 homophones (Ogura & Wang 2006).

4. Sherman lists 150 diatone pairs, but we left out three (*recast, repeat, upright*), because Sherman does not give historical description to them and their diatone status is not attested in contemporary references.

5. Phillips (1984) groups Sherman's disyllabic noun-verb diatones and oxytonic homophones on the basis of the prefix, and compares their average word frequencies. However, we cannot see any effect of the prefix on word frequency. Moreover, her calculation is inexact, because Sherman's data are sometimes missing in her data.

Acknowledgements

Mieko Ogura is supported by grants from the Human Frontier Science Program and the Ministry of Education, Culture, Sports, Science and Technology of Japan. William S-Y. Wang is supported by HKRGC-GRF number 14611615. We wish to thank Michiyo Takeda, and Akihiko Suzuki of Hitachi Medical Co. for their assistance in the data collection and the NIRS experiments. We are also thankful for two reviewers' valuable comments.

Data

CELEX Lexical Database of English (version 2.5). 1995. Centre for Lexical Information, Max Planck Institute for Psycholinguistics, Nijmegen.

LDC Japanese Lexicon. 1997. Linguistic Data Consortium, University of Pennsylvania, Philadelphia.

The Brooklyn-Geneva-Amsterdam-Helsinki Parsed Corpus of Old English. 2000.

The Penn-Helsinki Parsed Corpus of Middle English. 2000.

References

Baayen, R. Harald, Richard Piepenbrock & Léon Gulikers. 1995. The CELEX lexical database [webcelex]. Philadelphia, PA: University of Pennsylvania, Linguistic Data Consortium.

Binder, Jeffrey R., Rutvik H. Desai, William W. Graves & Lisa L. Conant. 2009. Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex* 19. 2767-2796.

Campbell, Alastair. 1959. *Old English grammar*. Oxford: Clarendon.

Carroll, John B, Peter Davies & Barry Richman. 1971. *American heritage word frequency book.* Boston: Houghton Mifflin.

Crepaldi, Davide, Manuela Berlingeri, Eraldo Paulesu & Claudio Luzzatti. 2011. A place for nouns and a place for verbs? A critical review of neurocognitive data on grammatical-class effects. *Brain & Language* 116. 33-49.

Fedorenko, Evelina G., Steven Thomas Piantadosi & Edward A. Gibson. 2012. The interaction of syntactic and lexical information sources in language processing: the case of the noun-verb ambiguity. *Journal of Cognitive Science* 13. 249-285.

Ferre i Cancho, Ramon & Solé, Richard V. 2003. Least effort and the origins of scaling in human language. *Proceedings for National Academy of Sciences* 100 788-791.

Fudge, Erik. 1984. *English Word-Stress.* London: Allen Unwin.

Fukui, Yuich, Yusaku Ajichi & Eiji Okada. 2003. Monte Carlo prediction of nearinfrared light propagation in realistic adult and neonatal head models. *Applied Optics* 42. 2881-2887.

Haiman, John (ed.). 1985. *Iconicity in syntax.* Amsterdam & Philadelphia: John Benjamins.

Hotta, Ryuichi. 2013. The diatonic stress shift in Modern English. *Studies in Modern English* 29. 1-20.

Jespersen, Otto. 1954. *A Modern English grammar on historical principles*, Part 1. London: Routledge.

Kempson, Ruth. 1980. Ambiguity and word meaning. In Sydney Greenbaum, Geoffrey Leech & Jan Svartvik (eds.), *Studies in English linguistics: for Randolph Quirk*, 7-16. London & New York: Longman.

Lass, Roger. 1992. Phonology and morphology. In Norman Blake (ed.), *The*

Cambridge history of the English language, volume II: 1066-1476, 23-155. Cambridge: Cambridge University Press.

Miller, George A. 1981. *Language and speech*. San Francisco: W. H. Freeman.

Moore, Samuel. 1967. *Historical outlines of English sounds and inflections.* revised by Albert H. Marckwardt. Ann Arbor: George Wahr.

Nakao, Toshio. 1978. *The prosodic phonology of Late Middle English*, Tokyo: Shinozaki Shorin.

Ogura, Mieko. forthcoming. *Language evolution as a complex adaptive system: a multidisciplinary approach to the history of English.* New York: Oxford University Press.

Ogura, Mieko & William S-Y. Wang. 2006. Ambiguity and language evolution: evidence of homophones and syllable number of words. *Studia Anglica Posnaniensia* 42. 3-30.

Ogura, Mieko & William S-Y. Wang. 2012a. The global organization of the English lexicon and its evolution. In Hans Sauer & Gaby Waxenberger (eds.), *English historical linguistics 2008, volume II: words, texts and genres*, 65-83. Amsterdam & Philadelphia: John Benjamins Company.

Ogura, Mieko & William S-Y. Wang. 2012b. Ambiguity resolution and evolution of word order. In Thomas C. Scott-Phillips, Monica Tamariz, Erica A Cartmill & James R Hurford (eds.), *The evolution of language*: *proceedings of 9th International Conference*, 274-281. Singapore: World Scientific.

Ogura, Mieko & William S-Y. Wang. 2014. Evolution of tense and aspect. In Erica A. Cartmill, Sean Roberts, Heidi Lyn & Hannah Cornish (eds.). 2014. *The evolution of language: proceedings of the 10th international conference*, 213-220. Singapore: World Scientific.

Ogura, Mieko & William S-Y. Wang. 2018. Evolution of homophones and syntactic categories noun and verb. In Christine Cuskley, Molly Flaherty, Hannah Little, Luke McCrohon, Andrea Ravignani & Tessa Verhoef (eds.), *Proceedings of the 12th international conference on the evolution of language*, 355-363. Torun: Nicolaus Copernicus University.

Phillips, Betty S. 1984. Word frequency and the actuation of sound change. *Language* 60. 320-342.

Piantadosi, Steven T. Harry Tily & Edward Gibson. 2012. The Communicative function of ambiguity in language. *Cognition* 122. 280-

291.

Sereno, Joan A. & Allard Jongman. 1995. Acoustic correlates of grammatical class. *Language and Speech* 38/1. 57-76.

Shapiro, Kevin & Alfonso Caramazza. 2003. The representation of grammatical categories in the brain. *Trends in Cognitive Sciences* 7/5. 201-206.

Sherman, Donald. 1975. Noun-verb stress alternation: an example of the lexical diffusion of sound change in English. *Linguistics* 159. 43-71.

Strang, Barbara M. H. 1980. The ecology of the English monosyllable. In Sydney Greenbaum, Geoffrey Leech & Jan Svartvik (eds.), *Studies in English linguistics: for Randolph Quirk*, 277-293. London & New York: Longman.

Tsigka, Stylliani, et al. 2014. Distinguishable neural correlates of verbs and nouns: a MEG study on homonyms. *Neuropsychologia* 54. 87-97.

Vonk, Jet M. J., Roel Jonkers & Loraine K. Obler. 2015. Semantic subcategories of nouns and verbs: A neurolinguistic review on healthy adults and patients with Alzheimer's disease. In Corine Astésano & Mélanie Jucla (eds.), *Neuropsycholinguistic perspectives on language cognition: essays in honour of Jean-Luc Nespoulous*, 61-74. London & New York: Psychology Press.

Wang, William S-Y. 2011. Ambiguity in language. *Korean Journal of Chinese Language and Literature* 1. 3-20.

Watanabe, Eiju et al. 1996. Non-invasive functional mapping with multi-channel near infra-red spectroscopic topography in humans. *Neuroscience Letters* 205. 41-44.

Yamashita, Yuichi et al. 1996. Near-infrared topographic measurement system: imaging of absorbers localized in scattering medium. *Review of Scientific Instruments* 67. 730-732.

Zempleni, Monika-Zita, Remco Renken, John C J Hoeks, Johannes M Hoogduin & Laurie A Stowe. 2007. Semantic ambiguity processing in sentence context: evidence from event-related fMRI. *NeuroImage* 34. 1270-1279.

Zipf, Kingsley. 1949. *Human behavior and the principle of least effort: an introduction to human ecology.* Cambridge, MA: Addison-Wesley.

Appendix: The Evolution of Diatones in the 16th, 17th, and 18th Century

word	diatone	n/v	OED source	OED first citation	homophone	O/P/D	word frequency
outlaw	16c	v	LOE	10..	16	D, P	2.7833
outlaw	16c	n	LOE	924	16	D, P	2.7833
rebel	16c	n	F	1340	16	D	3.0981
rebel	16c	v	F	1340	16	D	3.0981
record	16c	n	OF	a1300	16	D, O	115.97
record	16c	v	OF	a1225	16	D, O	115.97
convert	16c	n	convert (v)	1561	16	D	5.8238
convert	16c	v	OF	a1300	16	D	5.8238
desert	16c	n	OF	a1225	16	D	78.870
desert	16c	v	mod. F.	1539	16	D	78.870
incense	16c	n	OF	c1290	16	O, D	0.8155
incense	16c	v	F	1303	16	O, D	0.8155
present	16c	n	OF	a1225	16	D	137.90
present	16c	v	OF	c1290	16	D	137.90
refuse	16c	n	OF	c1440	16	O, D	8.9941
refuse	16c	v	F	13..	16	O, D	8.9941
accent	17b	n	Fr	1538	16	P, D	19.386
accent	17b	v	Fr	1530	16	P, D	19.386
collect	17b	n	F, L	a1225	16	D	31.686
collect	17b	v	OF, L	1573	16	D	31.686
compound	17b	n	compound (adj)	1530	16	D, O	58.832
compound	17b	v	OF	c1374	16	D, O	58.832
conduct	17b	n	OF or L	c1290	16	D, P	11.899
conduct	17b	v	F or L	c1400	16	D, P	11.899
contract	17b	n	OF	c1315	16	O, D	14.346
contract	17b	v	L	1530	16	O, D	14.346
convoy	17b	n	F	1500-20	16	D, P	0.4747
convoy	17b	v	F	1375	16	D, P	0.4747
object	17b	n	partly ppl a, partly L	c1380	16	D, P	117.35
object	17b	v	L	c1400	16	D, P	117.35
relapse	17b	n	relapse (v)	1533-4	16	O, D	0.0322
relapse	17b	v	L	1568	16	O, D	0.0322
torment	17b	n	OF, ONF	c1290	16	D	0.8642
torment	17b	v	OF	c1290	16	D	0.8642
abstract	17c	n	L	1528	16	D	2.7941
abstract	17c	v	abstract (ppl a)	1542	16	D	2.7941
cement	17c	n	OF	c1300	16	D, O, P	13.289
cement	17c	v	cement (n)	1340	16	D, O, P	13.289
compact1	17c	n	ppl a(1)<L	1601	17	D	7.2348
compact1	17c	v	compact (a)<L	1530	17	D	7.2348
confine	17c	n	F	c1400	16	D	0.6545
confine	17c	v	F	1523	16	D	0.6545
conflict	17c	n	L	c1430	16	D, P	16.476
conflict	17c	v	L	1432-50	16	D, P	16.476
conserve	17c	n	F	1393	16	O, D, P	2.5129
conserve	17c	v	F	c1374	16	O, D, P	2.5129
consort	17c	n	F	1419	16	D, P	0
consort	17c	v	consort (n1)	1588	16	D, P	0
contest	17c	n	contest (v)	1643	17	D, O	22.184
contest	17c	v	F	1579	17	D, O	22.184
converse	17c	n	converse (v)	1610	17	D, P, O	2.5234
converse	17c	v	F	1340	17	D, P, O	2.5234
convict	17c	n	convict (ppl a.)<L	1530-1	16	O*, D	0.1133
convict	17c	v	L	c1366	16	D	0.1133
essay	17c	n	OF	1597	16	D, O, P	4.5664
essay	17c	v	ASSAY (<L)	1483	16	D, O, P	4.5664
exile	17c	n	OF	1300	16	P, P, D	2.0964
exile	17c	v	OF	1330	16	P, D	2.0964
extract	17c	n	L	1549	16	O, D	3.4555
extract	17c	v	L	c1489	16	O, D	3.4555
ferment	17c	n	Fr	c1420	16	D	0.4827
ferment	17c	v	F	1398	16	D	0.4827
insult	17c	n	F	1603	17	D, O	2.0882
insult	17c	v	L	1570-6	17	D, O	2.0882
outcast	17c	n	out-<OE + cast(v)<ON	13..	16	D, P*, O*	0.4082
outcast	17c	v	out-<OE + cast(v)<ON	a1300	16	D	0.4082
project	17c	n	L	a1400-50	16	D	21.216
project	17c	v	L	c1477	16	D	21.216
subject	17c	n	OF	13..	16	D, P	112.47
subject	17c	v	OF or L	1382	16	D, P	112.47
transport	17c	n	transport (v)	1456	16	D, O	6.2411

transport	17c	v	F or L	c1374	16	D, O	6.2411
bombard	18b	n	OF	1393	16	P, O, D	0
bombard	18b	v	F	1598	16	P, O, D	0
compress	18b	n	F	1599	16	O, D	0.6516
compress	18b	v	OF	1398	16	O, D	0.6516
concert	18b	n	F	1665	17	O, D, P	9.5770
concert	18b	v	F	1598	17	O, D, P	9.5770
concrete	18b	n	L	1528-1725	17	O*, D	14.067
concrete	18b	v	concrete (a)<L	1635	17	D	14.067
confect	18b	n	med. L	1587	16	P, O, D	0
confect	18b	v	L	1545	16	P, O, D	0
contrast	18b	n	F	1597	16	O, P, D	22.787
contrast	18b	v	OF	1489	16	O, P, D	22.787
discord	18b	n	OF	1297	16	P*, P, D	0.4495
discord	18b	v	OF	a1300	16	P, D	0.4495
discount	18b	n	F	1622	17	O, P, D	1.9460
discount	18b	v	OF	1629	17	O, P, D	1.9460
export	18b	n	export (v)	1690	17	O, D	4.4660
export	18b	v	L	c1485	17	O, O*, D	4.4660
import	18b	n	import (v)	1588	16	O, D	4.2570
import	18b	v	L	c1430	16	O, D	4.2570
impress 1	18b	n	impress (v1)	1590	16	O, D	2.4433
impress 1	18b	v	L	c1374	16	O, D	2.4433
prelude	18b	n	F	1561	17	O, D	0.1668
prelude	18b	v	L	1655	17	O, D	0.1668
produce	18b	n	produce (v)	1699	17	D, O	100.37
produce	18b	v	L	1499	17	D, O	100.37
survey	18b	n	survey (v)	1535	16	O, D	9.6180
survey	18b	v	AF	1467-8	16	O, D	9.6180
undress	18b	n	un-<OE + dress (n)<dress (v)	1683	17	O, D	0.8570
undress	18b	v	un-<OE + dress (v)<OF	1596	17	O, D	0.8570
affix	18c	n	Fr	1612	17	O, D	0.2770
affix	18c	v	med. L	1533	17	O, D	0.2770
decrease	18c	n	OF	1383	16	O, D	3.3580
decrease	18c	v	OF	1382	16	O, D	3.3580
defile 1	18c	n	F	1685	18	O, D	0.4014
defile 1	18c	v	F	1705	18	O, D	0.4014
descant	18c	n	OF	c1380	16	O, P, D	0.2587
descant	18c	v	OF	c1510	16	O, P, D	0.2587
digest	18c	n	L	1387	16	O, P, D	2.0360
digest	18c	v	L	c1450	16	O, P, D	2.0360
increase	18c	n	increase (v)	c1374	16	O, D	43.379
increase	18c	v	AF	13..	16	O, D	43.379
inlay	18c	n	inlay (v)	1656	17	O, D	0
inlay	18c	v	in- <OE + lay (v)<OE	1596	17	O, D	0
outleap	18c	n	out- <OE + leap (n)<OE	c1250	16	O, D	0
outleap	18c	v	out- <OE+leap (v)<Com.Teut.	1600	16	O, D, O*	0
outwork	18c	n	out-<OE + work (n)<OE	c1615	17	P*, D, O*	0
outwork	18c	v	out- <OE + work(v)<OE	c1250	17	D	0
perfume	18c	n	F	1533	16	O, D	8.9841
perfume	18c	v	F	1538	16	O, D	8.9841
permit	18c	n	permit (v)	1714	18	O, P, D	12.290
permit	18c	v	L	1489	18	O, P, D	12.290
prefix	18c	n	mod. L	[1614] 1646	17	O, D	10.702
prefix	18c	v	OF	c1420	17	O, D	10.702
presage	18c	n	F	1390	16	P, O, D	0
presage	18c	v	F	1562	16	P, O, D	0
protest	18c	n	F	c1400	16	O, D	6.4360
protest	18c	v	F	1440	16	O, D	6.4360
purport	18c	n	AF	[1278] 1455	16	P, D	0.1127
purport	18c	v	AF	[1300] 1528	16	P, D	0.1127
regress	18c	n	L	c1375	16	P*,O*,O,D	0
regress	18c	v	L	1552	16	P*, O, D	0
reprint	18c	n	reprint (v)	1611	17	O*, D	0
reprint	18c	v	re- <L + print (v)<print (n)<OF	1551	17	D	0
surcharge	18c	n	surcharge (v)	1569	16	O, D	0
surcharge	18c	v	OF	1429	16	O, D	0
transfer	18c	n	transfer (v)	1674	17	P, D	7.9900
transfer	18c	v	F or L	1382	17	P*, P, D	7.9900
transverse	18c	n	L	1596	16	D	0.2062
transverse	18c	v	OF	1377	16	D	0.2062
uprise	18c	n	up-<OE+rise (n)<rise (v)<Com.Teut	a1300	16	O, D	0
uprise	18c	v	up- <OE+rise (v)<Com. Teut.	a1300	16	O, D	0