

# Corpus-based approaches and discourse analysis in relation to reduplication and repetition<sup>☆</sup>

Shih-ping Wang<sup>a,b</sup>

<sup>a</sup>*Department of Applied English, Ming Chuan University*

<sup>b</sup>*School of English Studies, University of Nottingham, University Park, Nottingham NG7 2RD, UK*

~ *The sound should be an echo to the sense. A. Pope*

---

## Abstract

Reduplication is important in language studies. Its linguistic form at the lexical level has long been explored in terms of various formalist theories. However, the linguistic function at other levels such as the discourse layer tends to be ignored. A reduplication corpus (ongoing compilation; 1687 items in total thus far) has been constructed as the baseline for an integrated approach to the interplay of various kinds of repetition in the use of language. The frequency of each token was calculated based on its occurrence in the British National Corpus (BNC). Then a wordlist with the top 102 items was proposed for related research topics such as frequency, percentage coverage, concordance, and collocation in terms of McCarthy's framework (1990 and later) using MonoConc Pro, WordSmith 4.0 and the SARA 3.2 software. The probability of collocation was calculated in terms of mutual information (MI). The higher the MI score, the more genuine the association between two items (Church and Hanks, 1990). A powerful search engine, Google, was further employed to locate relevant texts on websites for the analysis of reduplication from lexical to discourse levels. Both reduplication and repetition do play a significant role and exhibit extensively a certain language musicality in our everyday life.

© 2004 Elsevier B.V. All rights reserved.

**Keywords:** Corpus linguistics; British National Corpus (BNC); Mutual information (MI); Form and function; Reduplication; Discourse analysis; Advertising

---

<sup>☆</sup> Portions of this research were presented at the 2001 International Conference on the Application of English Teaching, Taipei (cf. Wang, 2001b), The Spring Meeting of the Linguistics Association of Great Britain (LAGB) 2002, Edge Hill, UK and 13th World Congress of Applied Linguistics (AILA) 2002, Singapore. This paper has benefited greatly from comments received during these presentations.

*E-mail addresses:* [spwang2005@yahoo.com.tw](mailto:spwang2005@yahoo.com.tw), [aexspw@nottingham.ac.uk](mailto:aexspw@nottingham.ac.uk).

## 1. Introduction

Corpus-based approaches (CBA) have been widely used to explore both written and spoken texts in recent years (Sinclair, 1997; McCarthy, 1998, and many others). Several advantages come from the use of CBA, for instance, the investigation of word usage, frequency, collocation and concordance (e.g., McEnery and Wilson, 2001; Scott, 2001; O’Keeffe and Farr, 2003). The fundamental features of corpus-based analysis include the following (Biber et al., 1998; Conrad, 1999):

- It is empirical, analyzing the actual patterns of use in natural texts.
- It utilizes a large and principled collection of natural texts as the basis for analysis.
- It makes extensive use of computers for analysis, using both automatic and interactive techniques.
- It depends on both quantitative and qualitative analytical techniques, especially functional interpretations of language use.

Given these methods, the aim of this research is to investigate reduplication and repetition at different levels (e.g., phonological, lexical, phrasal, discourse and even pragmatic<sup>1</sup>), and how they are used in contexts (*qualitative*). Relevant topics such as frequency, percentage coverage, concordance, collocation and mutual information are also explored (*quantitative*).

The research framework was mainly based on McCarthy’s lexical model, discourse analysis models (1990, 1991 and many later articles; see Appendix B) and his findings (1988). Both the *Language Knowledge Framework* (Celce-Murcia and Olshtain, 2000)<sup>2</sup> and the *Interconnected Dimensions of Grammar* (Celce-Murcia and Larsen-Freeman, 1999)<sup>3</sup> are used to reinforce the integration of form, meaning and function into the research of reduplication in different areas of language use (e.g., Leech, 1966, 1983; McCarthy and Carter, 1994; Cook, 2001; Scott, 2001).

### 1.1. Multiword units and reduplication

In their research into word usage, linguists have turned their attention to multiword units (MWUs) or formulaic sequences (FSs) (McCarthy, 1990; Wray, 2002; Spöttl and McCarthy, 2003).<sup>4</sup> Sinclair (1991, 1997) argues that the traditional priority given to the single word should yield to a much higher concern for MWUs, which are strings of words acting as a single lexical item with a single meaning (Schmitt, 2000; Wray, 2000; Grant and Bauer, 2004). Moon (1997: 43–44) proposes three significant criteria to distinguish multiword items from other types of strings: viz., institutionalization/lexicalization, fixedness and non-compositionality. She defines a multi-word item as a vocabulary item

<sup>1</sup> It is not the author’s intention to determine the borderline of discourse and pragmatics here.

<sup>2</sup> Please also refer to their more specific model, *Discourse Processing Framework* (p. 15).

<sup>3</sup> See also DeCarrico and Larsen-Freeman (2002).

<sup>4</sup> The names of MWUs are various: preformulated language, formulas, lexical phrases and multiword items (Nation and Meara, 2002; Moon, 1997; Wray, 2002: 8–9).

Table 1  
Examples for orthographic variations in the BNC

A	hiphop (2)	mishmash (24)	riffraff (2)	-----	zigzag (78)
B	hip-hop (43)	mish-mash (31)	riff-raff (43)	tit-for-tat (35)	zig-zag (102)
C	hip hop (97)	mish mash (2)	riff raff (6)	tit for tat (92)	zig zag (19)

Note: The numerals for each token indicate their frequency of occurrences in the BNC.

that includes a sequence of two or more words, semantically or syntactically constructing a meaningful, indivisible unit.

MWUs include compounds, phrasal verbs, idioms, fixed phrases and prefabs,<sup>5</sup> e.g., ‘first and foremost’ (248),<sup>6</sup> ‘wear and tear’ (153), ‘tit for tat’ (92) and the like (Moon, 1998; McCarthy, 1990; Wang, 2003a, 2003b). Different types of MWUs/FSs (such as compounds or fixed expressions) can be written as multiple orthographic words, hyphenated words or a single orthographic word, and this has implications for corpus searching and for transcription. The written forms can be different from person to person since there is no standardized spelling (Schmitt, 2000). For instance, the reduplicated forms or sound symbolism (SS) in Row A vary from those in Row B or C in Table 1. Their striking differences mainly lie in hyphenation and ‘space’:

It is worth noting that a corpus such as the British National Corpus (BNC) suggests that all the three forms of examples in Table 1 are grammatical in that they are used by educated writers of English.

As described above, this research uses quantitative and qualitative methods to explore reduplication and reduplicative fixed expressions in the form of binomials and trinomials at lexical, phrasal and discourse levels. Both corpus-based approaches and discourse analysis are adopted to deal with the phenomena of repetition and reduplication (hereafter R/R). A preliminary reduplication corpus (673 tokens for the pilot study) has been constructed (to date 1687 tokens in total). The frequency of occurrence for each token appearing in the BNC is established in order to come up with a list of the top 102 tokens as the baseline for research.

## 1.2. Research questions and problems encountered

The formal features of R/R have been extensively studied, but the reduplicative MWUs/FSs and sound symbolism (SS) have been neglected to a great extent (Ohala, 2001; Carter et al., 2001).<sup>7</sup> Previous studies rarely go beyond the interaction between phonological/lexical and discourse levels (cf. Conrad, 1999). Corpus-based approaches are therefore urgently needed to bridge the gap between formalism and functionalism. The research questions of the present study will, as a consequence, focus on the following topics to be justified in terms of integrated methods:<sup>8</sup>

<sup>5</sup> Freezes or fixed expressions include *irreversible conjoined phrases* and *fixed reduplicatives* (Pinker and Birdsong, 1979). For details, see Carter (1998) and Moon (1998). MWU includes compound words, phrasal verbs, fixed phrases, idioms and proverbs (Schmitt, 2000: 99–100).

<sup>6</sup> The Arabic numeral indicates frequency of occurrence in the British National Corpus (BNC).

<sup>7</sup> The association between sound and meaning has been regarded with suspicion by mainstream linguists.

<sup>8</sup> Concerning some similar research questions, interested readers are also referred to the related references in parentheses.

- Reduplication has been over-emphasized at the lexical level, at the expense of discourse or pragmatics. Therefore R/R will be discussed at different levels; the interaction of various layers will be given greater attention, especially the phonemic/phonological and discourse levels (e.g., Tannen, 1989).
- Examples illustrating the interplay between exact repetition and reiteration will be presented to confirm McCarthy's findings (McCarthy and Carter, 1994).
- Most research into reduplication is form-oriented, not functional or pragmatic.<sup>9</sup> Function is thus addressed and integrated into the R/R analysis (cf. Moravcsik and Wheatly, 1998: 1–7; Hoey, 1991).
- The frequency and probability of collocation for MWUs or SS reduplication will be further calculated in terms of mutual information (MI) (Church and Hanks, 1990).

The three dimensions of morphosyntax (*form*), semantics (*meaning*) and pragmatics (*use*) are of equal importance (Widdowson, 1989; Celce-Murcia and Larsen-Freeman, 1999). Their integration will be of primary interest to the current topic; therefore, this research is not confined to using one theory to the exclusion of another. It aims to deal with R/R at different levels, and to bridge the gap between formalism and functionalism (cf. Darnell et al., 1998; Moravcsik and Wheatly, 1998, *Introduction*) in terms of an integrated methodology. Everyday examples from phonemic, lexical to discourse levels are provided to further strengthen this study.

## 2. Literature review

### 2.1. Research into reduplication/repetition (R/R)

#### 2.1.1. The significance of R/R

Mainstream linguistics has conventionally concentrated on formal features in terms of classic generative approaches during the several past decades (Marantz, 1982; Steriade, 1988; Katamba, 1993, among others). The current topic tends to be neglected because of its structural variations (cf. Jakobson and Waugh, 1979). In addition, using repetition, considered 'bad style', is discouraged in native learners of English (Cook, 1989) and is often regarded negatively (Tannen, 1989; McCarthy and Carter, 1994).

However, it is not sufficient to say that the traditional beliefs consider R/R theoretically awkward or irrelevant. Some linguists argue that R/R deserves more attention in different genres (Cooper and Ross, 1975; Johnstone, 1987; McCarthy, 1988; Fenk-Oczlon, 1989; Tannen, 1989; Birdsong, 1995; Landsberg, 1995a, 1995b; Minkova, 2000). Scholars such as these have frequently emphasized the importance of R/R, which can be used not only "for reinforcement, generally with emotional emphasis", but also for intellectual purposes (Lausberg, 1960/1998).<sup>10</sup> Human beings generally use R/R from childhood onward (Jespersen, 1964; Ingram, 1974; Ferguson, 1983),<sup>11</sup> and therefore the reduplication

<sup>9</sup> Formal pragmatics is still used for the current research. Traditional pragmatics is ignored here due to limited space. Interested readers are also referred to Kadmon (2001).

<sup>10</sup> Thanks go to an anonymous reviewer who provided the reference.

<sup>11</sup> Consider that reduplication and repetition are highly used in children language acquisition. For more arguments, see Ingram (1974) and Ferguson (1983).

phenomenon is universal and significant for human communication. As proposed by McCarthy and Carter (1994), all conversations include repetition, and it occurs in all types of discourses. Repetition “is the central linguistic meaning-making strategy, a limitless resource for individual creativity and interpersonal involvement” (Tannen, 1989: 97). It is “a good example of speakers talking, creating” (Carter and McCarthy, 2004: 65). Garner (2002: 168–176) stresses that the fundamental principles of legal writing include various forms of repetition. That is, repetition is used to avoid ambiguity in legal English.<sup>12</sup> Finnegan (1977: 90) even proposes that repetition is definitely the most salient feature of poetry. As for areas of linguistic research, R/R spreads over a wide range of sub-disciplines:

- the study of nursery language and child language acquisition (Jespersen, 1964; Ingram, 1974; Schwartz et al., 1980; Fee and Ingram, 1982; Tannen, 1989: Chapter 3);
- research in general communication (Jespersen, 1964; Ferguson, 1983);
- the study of language play, language learning and mnemonics (Leech, 1966; Nation, 1982; McCarthy, 1990; Cook, 2000);<sup>13</sup>
- discourse analysis and lexical cohesion (McCarthy, 1991; Halliday, 1994; Cook, 2001);
- the study of back-channel responses (McCarthy, 2002; McCarthy and Walsh, 2003);<sup>14</sup>
- the study of language and affect (Leech, 1966; Hinton et al., 1994);
- the study of media language, e.g., advertisements and brand-names, slogans and headlines (Wales et al., 1958; Leech, 1966; Goddard, 1998; Beard, 2000; Cook, 2001);
- legal English and legal writing (Garner, 2002; Williams, 2004);
- alliteration and rhyming patterns in creative literature, poetry, lyrics, etc. (Pomorska et al., 1987; Waugh, 1987; Tannen, 1989; Carter, 1998; McRae, 1998; Yip, 2000).

### 2.1.2. *Different levels of language model*

Language can be classified into seven levels, which in turn can be grouped into three broad layers (cf. Jeffries, 1998: 5; see Table 2). Traditionally, reduplication is more emphasized at levels [1–3], while it seems to have long been ignored at the discourse level (layer [7]). In fact, levels [1–6] can also be explored at the contextual or discourse level in order to observe their semantic or pragmatic relationships. In general, R/R involves several layers such as lexical, phonological, syntactic and discoursal, including “phonemes, intonational and rhythmic patterns, words, idioms, phrases, sentences or discourse structures” (McCarthy and Carter, 1994: 144–147).

### 2.1.3. *The differentiation of R/R at different levels*

Linguistic terms relating to R/R indicate a great variety in terminology and scope. Thun (1963) proposes three criteria to distinguish repetition from reduplication using ‘pretty, pretty’ (6) and ‘pretty-pretty’ (1):<sup>15</sup> (1) a phonetic/prosodic difference, e.g., ‘*prétty-pretty*’;

<sup>12</sup> Thanks go to Christopher Williams (p.c., Feb. 11, 2004) who emphasizes that the frequent repetition of particular words, expression and syntactic structures are used in legal English to avoid ambiguity and redundant arguments.

<sup>13</sup> Seven repetitions are enough for most people to memorize a word (Nation, 1982). Repetition seems most effective if it begins very shortly after the initial learning.

<sup>14</sup> Consider the back-channels (‘discourse markers’) ‘uh-huh’, ‘ya-ya’ or ‘okay-okay’ in our daily conversation.

<sup>15</sup> Note: ‘pretty pretty’ (4).

Table 2

Different levels of language model (cf. Jeffries, 1998: 5). Revised and reproduced with written permission from the author and Palgrave Macmillan.

Units at each level			Levels of language
1. Sounds which combine to make	Phonology	}	LEXICAL LEVEL
2. Morphemes, which combine to make	Morphology		
3. Words, which combine to make	Lexis		
4. Phrases, which combine to make	Syntax	}	SYNTACTICAL LEVEL
5. Clauses, which combine to make	Syntax		
6. Sentences, which combine to make	Syntax		
7. Texts, which combine to make	Discourse	→	CONTEXTUAL LEVEL

(2) a morphological criterion, e.g., the possibility of plural, ‘pretty-pretties’; (3) a semantic difference.<sup>16</sup> Different terms are used to express similar phenomena, and some kinds of repetition can be distinguished in terms of different linguistic levels (Persson, 1974):

- lexical repetition
  - 1) I glance eastward and recognize the old, old (54)<sup>17</sup> view.
  - 2) Last but not least (59), reproductions were cheap and plentiful.
- syntactic repetition
  - 3) God he (94) knows what havoc I shall make.
  - 4) “Did you tell him that I ordered it?”–“Yes, I did.”
- semantic repetition
  - 5) They deceived and hoodwinked all of us.

The common denominator of all the above examples lies in the basic principle that *a sound or a concept is repeated in one form or another*. However, there is still a fundamental difference between reduplication and repetition:

- reduplication (lexical level) vs. repetition (syntactic level)
  - 6) They just let you go on buying on the never-never (19).
  - 7) Your Daddy learnt never, never to cry in the dark (113).

The hyphenated reduplication ‘never-never’ (19 occurrences; Noun/Adjective) is used at the lexical level, but the adverbial repetition ‘never, never’ with greater frequency of occurrences (113 occurrences) appears at the syntactic level.

#### 2.1.4. Relationship between R/R and SS

The naming and classification of the general terms R/R and SS are the scene of a tug of war among scholars. The relevant terms can be classified into at least twenty-four types (Wang, 2001b), revealing that ‘reduplication’ appears in more than 25% of the given entries. This implies that there is a strong tendency for R/R to be closely associated with SS.

<sup>16</sup> The repeated adjectives ‘pretty, pretty’ preserve their basic meaning, but the primary aim of reduplicative ‘pretty-pretty’ is overdone and becomes derogatory.

<sup>17</sup> The numeral represents the frequency of occurrence in the BNC.

SS is the relationship between the sound of an utterance and its meaning (Short, 1996; Carter et al., 2001; Ohala, 2001). Hinton et al. (1994) outline a typology of SS consisting of four categories: (1) *corporeal*, e.g., vocalizations and comic strips with visual effects; (2) *imitative*, e.g., onomatopoeia and sound-movement rhyming; (3) *synaesthetic*, e.g., diminutives and size symbolism, and (4) *conventional*, e.g., phonaesthesia, blending and creation of names for commercial products. Onomatopoeia appears with two semantic functions: ‘imitation’ and ‘name-making’. Diminutives such as ‘teeny-weeny’ exhibit sound reduplication as well. It is clear that both R/R and SS overlap to some extent.

#### 2.1.5. R/R and metaphoric extension

Lausberg (1960/1998) proposes that repetition has the informative and reinforcing functions. Moon (1998) classifies metaphors, including pure idioms and reflecting degrees of transparency, into three types, i.e. transparent, semi-transparent and opaque metaphors. Traditionally, metaphor is taken as a figure of rhetoric. However, Lakoff and Johnson (1980: 127–128) consider that metaphor is all-pervasive in our everyday life: witness the title of their work, *Metaphors We Live By*. They not only shed new light on metaphor, but also proposed, in relation to reduplication, “MORE OF FORM making MORE OF CONTENT”. The most representative devices highlighting ‘repetition’ consist of the following essential features, at different levels, along with synonymous or contrastive examples (Lakoff and Johnson, 1980: 128; cf. McCarthy, 1990; Carter, 1998; Regier, 1998; Green, 2000; Grant and Bauer, 2004).<sup>18</sup>

- *Reduplication makes a noun become a plural or collective:*

e.g., “She sat at the table, eating **bread** and **butter**.” (*food*)

“This **fund** raising fair is our **bread** and **butter** and it is top quality.” (*income*)

“Police have been conducting **house** to **house** enquiries at **homes** nearby.”

- *Reduplication is used to make a verb indicating ‘continuation or completion’:*

e.g., “There is a prophecy of Isaiah ...: You may **hear** and **hear**, but you will

never understand; you may **look** and **look**, but you will never **see**.”

(‘hear and hear’: to listen very carefully; ‘look and look’: to observe carefully)

- *Reduplication is applied to reinforce an adjective for intensification or increase:*

e.g., “Well we lived there er yes the **whole**, **whole** time, **all** our married life.”

“They were amazed at this **big**, **big** voice coming out of this **tiny** girl.”

“What happens to economic and **monetary** policy...**budgetary** policy...

which are the substantial **bread-and-butter** issues of our **politics**?”

(‘**bread-and-butter**’: adj., very important or the most basic).

<sup>18</sup> The features are defined by Lakoff and Johnson (1980: 128) with examples extracted from the BNC.

Table 3

The simplified structure of English reduplication

	Alternation	Examples
(1) <b>Full reduplication</b>		
V.V	exact copy	aye aye (95)
CV.CV		dodo (53), lulu (53), tutu (34), wee wee (37)
CVC.CVC		Tartar (39), dum dum (11), dum-dum (5)
(2) <b>Partial reduplication</b>		
C <sub>1</sub> V.C <sub>2</sub> V	C <sub>1</sub> ~ C <sub>2</sub>	polo (565), hi-fi (228), voodoo (76),
CV <sub>1</sub> C.CV <sub>2</sub> C	V <sub>1</sub> ~ V <sub>2</sub>	criss-cross (55), sing-song (77), zig-zag (102)
C <sub>1</sub> VC <sup>n</sup> .C <sub>2</sub> VC <sup>n</sup> (n > 0)	C <sub>1</sub> ~ C <sub>2</sub>	hubbub (89), ragbag (31)
C <sub>1</sub> V.C <sub>2</sub> V. C <sub>3</sub> V.C <sub>2</sub> V	C <sub>1</sub> ~ C <sub>3</sub>	charivari (32), willy-nilly (69)
(3) <b>Beyond word level</b>		
CVC ~ CVC	exact copy	so and so (215), such and such (130)
C <sub>1</sub> VC ~ C <sub>2</sub> VC	C <sub>1</sub> ~ C <sub>2</sub>	wear and tear (153), Queen of Mean (0)
CV <sub>1</sub> C ~ CV <sub>2</sub> C	V <sub>1</sub> ~ V <sub>2</sub>	this and that (212), tit for tat (92)
...	...	...

cf. <http://ccat.sas.upenn.edu/~haroldfs/echoword/echodata.html>.

- *Reduplication turns something small to diminutive:*

e.g., “Just a **little** pocket, a **little tiny, tiny** torch with a very **strong** beam.”

“A **weeny weeny** purse”; “Just a **bit, weeny, weeny weeny**”

As indicated above, R/R is also related to diminutive, e.g., ‘teeny-weeny’, which in turn manifests SS (Hinton et al., 1994; Jurafsky, 1996; Regier, 1998). A diminutive reduplication in SS also involves metaphoric extension. It is commonly accepted that the high front vowel /i/ is symbolic of smallness (Sapir, 1929), particularly indicating “what is small, weak, insignificant, refined or dainty” (Jespersen, 1933/1964). The vowels /u, o, a/ are more commonly related to something larger. The vowel sounds in ‘click-clack’, ‘tittle-tattle’ or ‘clink/clank’ have similar alternations (little-large) with /i-a/ or /i-æ/ ablaut.

## 2.2. Form, formalism and problems

### 2.2.1. Word-formation and ablaut for reduplication

Reduplicated word-formation varies in English, for example /i-æ/ (riprap) and /i-o/ (ping-pong), but it can be simply formulated as shown in Table 3.

In fact, reduplication forms can be further categorized into four broad types and sixteen subtypes in total as summarized below (Wescott, 1980; see Appendix A for details; Davis and Hammond, 1995; Yip, 2000):

- Type A: six subtypes of intra-syllabic homophonation [palindrome], e.g., gig, gag;<sup>19</sup>

<sup>19</sup> As S. Davis (p.c., 2002) points out, “there are no monosyllabic palindrome words in English that begin with consonant clusters such as \*flilf, \*prarp, \*spops, etc. The only exceptions are some st-plural words like “states” and “stats”...” Wescott (1980: 364–366) proposes 10 exceptional examples and explains why. Some of the examples undergo infixion and violate phonotactical rule in English palindromes.



- Type B: three subtypes of simple extra-syllabic homophonation, e.g., bamboo, dim-wit;
- Type C: four subtypes of compound extra-syllabic homophonation, e.g., flimflam;
- Type D: three subtypes of combinations of intra-syllabic and extra-syllabic repetitions, e.g., boob-tube, poppy-cock.

### 2.2.2. Formalism and problems

English reduplication in the sense to be stressed here is sometimes called ‘partial reduplication’ because it involves consonant ablaut or vowel alternation. It is common but not so productive and regular in sound-symbolic forms which classic generative linguists have failed to formalize into universal rules (Katamba, 1993). The alternations for each type of formation, including consonant changes, can be simply formulated along the following lines (cf. Marchand, 1969; Marantz, 1982; Kimenyi, 1989):

- *Onset alternation*:
  - a. ‘h-C’ alternation (C: consonant): handy-dandy/hobnob/hurly-burly
  - b. ‘C-w’ alternation: bow-wow/popsy-wopsy/teeny-weeny/wishwash
  - c. other alternation: fuddy-duddy/killer-diller/loco-foco/love-dovey/mumbo-jumbo/pokemoke/ragtag/ram-jam/super-duper/tagrag/titbit/
- *Vowel alternation* (orthographical):
  - a. ‘i-a’ alternation: chitchat/drizzle-drazzle/rip-rap/zigzag
  - b. ‘i-o’ alternation: crisscross/ding-dong/flipflop
  - c. others: gewgaw/seesaw/shuffle-shuffle
- *Rhyme with diminutive suffix* (e.g., in nursery rhymes):
  - a. ‘-ie’: boogie-woogie/heebie-jeebies/hootchie-kootchie/peepie-creepie
  - b. ‘-y’: flibberty-gibberty/ fuddy-duddy/hoity-toity/love-dovey

Assuming that the root is the first syllable of the reduplicated form, the derivation of a form such as *zigzag* basically undergoes two stages, 1) syllable copy; 2) vowel ablaut:

- 1) A syllable template is fully reduplicated and linked to its C/V slots (CVC→CVC.CVC; i.e. *zig* → *zig.zig*).
- 2) A vowel change motivated by ablaut occurs from /i/ to /a/ ([æ]) for the second syllable (*zig.zig* → *zig-zag* or *zigzag*).

If two pseudo-morphemes are put together, their rhymes (nuclear + coda) are the same, e.g., *hobnob* and *fuddy-duddy*, and fundamentally undergo two stages as well. For example, *hobnob* is considered as being within the group of /h~n/ alternation:

- 1) A syllable template is fully reduplicated and linked to its C/V slots (CVC→CVC.CVC; i.e. *nob*→*nob.nob*).
- 2) An onset alternation motivated by /h~n/ ablaut occurs from /n/ back to /h/ for the first syllable (*nob.nob* → *hobnob*).

However, both of the processes above can only be formulated for ‘regular’ examples to some extent. They cannot be always applied, especially for less regular patterns such as ‘fancy-shmancy’ and ‘hullabaloo’. Such borderline cases are problematic with obvious compounds containing bases that are non-existent as independent words. Thus, two thousand reduplication samples (Thun, 1963) still challenge us with abundant unresolved problems (Jakobson and Waugh, 1979). They are primarily formulated as follows:

- It cannot be predicted that each pattern always undergoes the same alternation.
- Since the variations are not deterministic, the formulated rules should be modified based on different individual cases.
- Formal solutions trying to explain English reduplication phenomena are not always deterministic and are more likely to be probabilistic (Carter and McCarthy, 1999; Bod et al., 2003). Deterministic solutions could be misleading because frequent exceptions are found (e.g., seesaw, shuffle-shuffle).<sup>20</sup>
- It is also accepted that not all reduplications are meaningful or have pragmatic outcomes.<sup>21</sup>

As regards the ‘h-C’ alternation above, Kimenyi (1989) claims that if the leftmost onset changes, it *always* changes into an ‘h’ sound. It is questionable whether such a strong statement can be made, given the many variations. The statement can best be expressed as a tendency, not a rule, to be explored based on the probabilistic occurrence of each item in a corpus. Therefore, probabilistic terms and corpus evidence are needed to reinforce the theory and make up for any weakness in this regard.

### 2.3. Discourse, pragmatics and R/R functions

#### 2.3.1. Relationship between discourse and pragmatics

Pragmatics and discourse analysis (DA) explore language’s relation to the contextual background characteristics. Both have some essential features in common, viz., *context*, *text* and *function* (Cook, 1989; Cutting, 2002):

- Both approaches mainly investigate the meaning of words in context or in interaction and how interactors communicate more information than the words they use.
- Both observe language use (*discourse*), and pieces of spoken/written discourse (*text*), focusing on how parts of language become significant and unified for their users, i.e. *coherence* in discourse or *relevance* in pragmatics.
- Both are concerned with function, the speakers’ short-term aims in speaking, and long-term purposes in interacting orally.

However, DA is different from pragmatics in its stress on the structure of text.<sup>22</sup> DA focuses on the features connecting language with the contexts where it is used. In the case of spoken language, this includes everyday events such as telephone calls, conversations,

<sup>20</sup> Interested readers are also referred to Optimality Theory (Minkova, 2000; Yip, 2000).

<sup>21</sup> This is pointed out by Ronald Carter and an anonymous reviewer.

<sup>22</sup> “Pragmatics takes a socio-cultural perspective on language usage, examining the way the principles of social behavior are expressed is determined by the social distance between speakers” (Cutting, 2002: 3).

interviews and the like. As for written texts, linguists are more concerned with how texts are put together coherently, and how they form patterns (McCarthy, 1990, 1991).

In addition, DA has concerned itself with lexical cohesion and lexical signalling. The most direct form of lexical cohesion is the *repetition* of a lexical item (Halliday, 1994: 330–331; Cutting, 2002). Simple repetition is the simplest type of lexical relation (Hoey, 1991). Repetition is also a discourse device used to deal with topic negotiation: different speakers repeat each other's words to push their own topic forward (McCarthy, 1998). What all kinds of repetition links have in common is that they allow a speaker or writer to say something new by replacing existing information.

Tannen (1987, 1989) is concerned with synchronic, diachronic and *visual repetitions*. She proposes that repetition has a central role in how discourse is created. Three strategies primarily involving sound are also based on repetition, including (1) *rhythm* taken as a musical score or basic musical performance in conversation; (2) *patterns* based on repetition and variation of phonemes, morphemes, words, word collocations and longer sequences of discourse; (3) *style figures of speech*. Many of these are repetitive figures<sup>23</sup> as well.

### 2.3.2. Functional classification of reduplication

The function of repetition can be rhetorical and cohesive (cf. Károly, 2002: 98), acting as a mechanism for assimilating the new to the old. Cook (1989) elaborated the following categories for the construction of lexical chains of connected words in running discourse: 1) *repetition* (exact copy), sometimes inappropriate; 2) *elegant repetition*, where synonyms or more general words are preferred; 3) *referring expressions*, e.g., using a pronoun, 'this and that' and 'here and there' to refer; 4) *parallelism*, echoing forms used to repeat a sound or to create a grammatical or semantic parallelism, as in rhyme, rhythm, and the sound effects of verse.

Content can be reiterated in paraphrase form or alternative lexical forms (near synonyms). Meaning becomes fixed in context under the circumstances created by a speaker or a writer who changes the lexical choices. Repetition plays a specific role in the negotiation of lexical meaning among speakers. It is basically categorized into three types:

- 1) exact repetition (e.g., sees → sees)
- 2) reiteration (e.g., sees → sights)
- 3) relexicalization (e.g., sees → glimpses)<sup>24</sup>
  - recasting the same meanings in different words
  - using equivalents (synonyms) or superordinates.

In addition, interplay between exact repetition and reiteration will occur in discourse (McCarthy, 1988, 1990: 97; McCarthy and Carter, 1994).<sup>25</sup>

<sup>23</sup> A typical example is given in John F. Kennedy's words: "Ask not what your country can do for you. Ask what you can do for your country." There are 17 words in all, but only 'not' isn't repeated.

<sup>24</sup> Examples are provided by Michael McCarthy, personal communication, August 2001.

<sup>25</sup> For different classifications, see Károly (2002: 104).

### 2.3.3. Applications in everyday examples

SS reduplication in daily communication is discussed in the works of many linguists (Leech, 1966; McCarthy, 1990; Cook, 1989, 2000; McCarthy, 1998, 2000). It is routinely employed in advertisements, newspapers, election slogans and the like. SS is also often used for brand naming.<sup>26</sup> In the following six categories SS reduplication is commonly used:

- 1) Baby talk, children songs, lyrics, poetry, and prayer;
- 2) Second language pedagogy, phonics, especially for children's English;
- 3) Language games, tongue twisters, comics and cartoons;
- 4) Advertisements, branding, political slogans;
- 5) Headlines (for any message or newspaper);
- 6) Political and ideological rhetoric.

A typical case, the tongue twister, exhibits massive reduplication in terms of onset-repetition and coda repetition (cf. Kirshenblatt-Gimblett, 1976; Sanches and Kirshenblatt-Gimblett, 1976: 79). For instance, the analysis of the consonantal ablaut in “the blanket like a lipstick” illustrates a striking picture of musical interplay between stops and /l/ (Jakobson and Waugh, 1979: 218):

blanket	bl	n	kt
like	l		k
lipstick	l	p	stk

It is noteworthy that consonantal reduplicated patterns build on the interaction of parallel repetitions between onsets (/l/) and codas (/kt/~/k/~/stk/).

In advertisements, “...repetition is perhaps the major factor in making a slogan a byword” (Wales et al., 1958: 202; Halliday, 1994). Thus, R/R is often used for the sake of advertising effects. A typical example below, using ‘phonological repetition’ (Tannen, 1989: 77), illustrates how it is used in a car advertisement (Wales et al., 1958: 203):

When Better Automobiles Are Built, Buick Will Build Them					
W-n	B- t	t-m-b-l	B-lt, B-k	W- l	B-l d th-m

The repetition of sounds is employed again and again in this slogan with nine words only. Once the consonants are extracted, this slogan exhibits prominent features of repeated bilabial and alveolar sounds:

- Topic: **Buick** (B: specification of Automobiles)
- **B**: **Better, Built, Buick and Build** (B: onset-repetition)
- **ui**: **Built, Buick and Build** (ui: ‘quasi-nucleus/visual repetition’)<sup>(27)</sup>
- **W...B** pattern: **W**hen **B**etter...**W**ill **B**uild (/W...B/: bilabial onset-repetition at syntactical level)
- Pronoun: Automobiles = Them (Pronoun: grammatical cohesion)

<sup>26</sup> SS is helpful to decode branding language (<http://www.lexicon-branding.com/techniq/sound.htm>).

<sup>27</sup> The spelling ‘ui’ for **Built**, **Buick** and **Build** is orthographic (visual). The ‘ui’ sound /jul/ in Buick is different from that of the other two tokens, //l/, but still it has advertising effect.

Company name	McVITIE'S Original
Brand name	<b>H o b n o b s</b> .....
Slogan	<b>ONE <u>NIBBLE</u> AND YOU'RE <u>NOBBLED</u>.</b> <u>n</u> obly, <u>o</u> aty <u>b</u> iscuits

Fig. 1. 'Hobnob' advertisement.

The distinguishing attribute of this slogan is that it repeats the /b/-sound four times in terms of *alliteration*, i.e. onset-repetition, to stress the better quality of its product.

One more characteristic of metaphoric expression is the pattern of bilabial onset-repetition '**W...B**' used to highlight: 'We're Buick—We're Better.' A similar slogan, *One Nibble and You're Nobbled!*, integrates a *h-C* alternation and /i-a/ alternation for a cookie brand name, *Hobnobs*, (<http://shop.store.yahoo.com/dphouse/orhob.html>) (Fig. 1):

The slogan undergoes /i-a/ ablaut for both 'nibble' and 'nobbled'; they in turn reflect their sound reduplication back to the brand name 'Hobnob'. This alternation, a type of *Conventional Sound Symbolism*, occurs at the syntactic level and can be used to demonstrate lexical cohesion in terms of SS reduplication. As Hoey (1991) proposes, the *repetition link triangle* is applied to contextually connect one of the previous occurrences of a repetition link with all earlier occurrences (Fig. 2).

The triangle exhibits its visual repetition and the lexical cohesion contextually in terms of both lexical and sentential sound repetition. Its potential advertising effects are obtained by integrating into the discourse level (1) *synesthetic SS* (**Category 3**: representing visual properties) and (2) *conventional SS* (**Category 4**: brand-naming, with sound and meaning linked automatically).

### 3. Method and procedure of the present study

#### 3.1. Data and analytical tools

The data for the present study include the author's own reduplication corpus and the BNC. The author's data have been gathered from research papers, books, dictionaries,

- **Hobnobs** (rhyme, Type C.3 in Table 12)
- **Hob-** (SS: sound of gorging)
- **-nob** (3 key sounds also for nobly, oaty biscuits)
- **nibble...nobbled** (/i-o/ ablaut, Type C.2, at syntactic level)
- **nobly, oaty biscuits** (3 onsets from -nob)

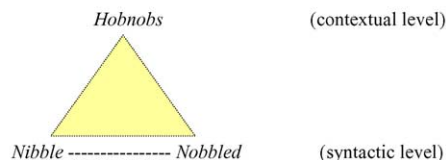


Fig. 2. Triangle for repetition link.

websites, newspapers, advertisements, slogans, etc. since 1986. The ongoing collection has undergone two stages (to date 1687 tokens, see the following section for discussion). The criteria for the establishment of the reduplication corpus are mainly based on form:

- The form of each token should be *reduplicated* in various types as outlined earlier.
- A token may be a reduplicative binomial/trinomial expression, SS, onomatopoeia and the like.

The reduplication corpus was first constructed and then each token was searched for its frequency of occurrence in the BNC, which is chosen because of its prominence in current corpus-based work (e.g., Biber et al., 1998; Scott, 2001); also, it is relatively user-friendly.<sup>28</sup> Its basic features are the following:

- The major point of reference is British English.
- It includes 100,106,008 words (4124 texts from 90% written, and 10% spoken sources).
- The texts are well-balanced for academic, educational and commercial aims.
- It is a demographical sampling with a wide range of genres from both spoken and written English, including one of the largest collections of spoken data ever available.
- Its context is contemporary, and it contains multiple examples of relatively rare items.

(Rundell, 1995; Aston and Burnard, 1998; Kennedy, 1998)

The Google search engine (<http://www.google.com>) was extensively used on the Internet to save search time and reinforce the analysis of texts. Its benefits are manifold:

- Its search covers a wide range of data (more than 1 billion URLs), including 16 categories.
- Google prioritizes results based on the individual search relevant to the query.
- Its Image Search is the most comprehensive on the Web.
- PDF format: Google finds high-quality information stored in PDF.
- The samples searched are genuine and vivid, consisting of figures, pictures, and tables, which a general corpus such as the BNC does not exhibit. For example, an original advertisement with any picture sometimes is more efficient and illuminating than one with text only.

Finally, the concordancing software packages, MonoConc Pro and WordSmith 4.0, were employed to generate frequency lists, concordances and collocation information. The built-in SARA software in the BNC was used to calculate mutual information scores for the probability of collocation. The following are the main areas to be explored in order to show the pervasive use of reduplication in day-to-day discourse:

- Author's own corpus: data from language play, lyrics, songs, poetry, newspapers, daily conversation, dictionaries, websites, etc.;

<sup>28</sup> Compiled by Oxford University Press, 1991–1994, the BNC is easily available on CD-ROM with the SARA software package and is appropriate for this current research, which is orientated towards British English (<http://info.ox.ac.uk/bnc/what/index.html>).

- The BNC, using SARA 3.2 search for the frequency of occurrence for each token;
- General websites, for surfing and downloading examples of reduplication.

As to instruments, Google, MonoConc Pro, WordSmith 4.0 and Sara-32 were used for quantified results, e.g., frequency, wordlist, percentage of the total words in the corpus, concordance and collocation.

### 3.2. Integrated framework

McCarthy's framework (mainly 1990) for English vocabulary is structured under five topics: *words* → *lexical relations* → *the mental lexicon* → *vocabulary in use* → *vocabulary as a learning corpus*. They can be connected with each other in a flowchart and as a framework for analysis of language at the lexical, syntactical and discourse levels (see [Appendix B](#)). In addition, Celce-Murcia and Olshtain's *Language Knowledge Framework* or more specific *Discourse Processing Framework* (2000: 14–15) can be used as a supplementary instrument to illustrate the relations between discourse and pragmatics (for details, see [Appendix C](#)). As mentioned earlier, "one approach should not be taken to the exclusion of others" (DeCarrico and Larsen-Freeman, 2002). An integrated framework therefore is required to provide a middle ground. The three dimensions of form, meaning and use are equally important for this research (Celce-Murcia and Larsen-Freeman, *Interconnected Dimensions of Grammar*, 1999; for details, see [Appendix D](#)).

### 3.3. Procedure

The qualitative analysis is mainly based on McCarthy's (1990) framework and various discourse analysis models. The quantitative analysis is founded on corpus approaches. In terms of the interactions of both analyses, *Linguistic Form* (Levels 1–6) and *Linguistic Function* (Level 7) in [Table 2](#) will be distinguished. The initial step is to collect and analyze data, and then create the first compiled wordlist (1) according to the following principles:

- All tokens should be reduplicated (see [Table 3](#));
- All tokens undergo BNC simple search for frequency of occurrence;
- All types include SS/non-SS reduplication, fixed expressions and idioms.

The basic procedure is a simple search of each token for its frequency of occurrence in the BNC. The next step is choosing the top 100 tokens to create wordlist 2. Example texts can then be found and extracted from the BNC and Google. The flowchart for this procedure is given in [Fig. 3](#).

To take just two illustrative examples, consider [Fig. 4](#), downloaded from a website (corresponding to *Category 1*, see [Section 2.1.4](#)):

The SS reduplication 'pitter patter' collocates with RAINDROPS, implying the rain falls *pit-a-pat* with many quick light beats. It is more easily understood by using sounds with the help of a picture. A further example of child-directed language is "tic-tac-toe", the name of a game (corresponding to *Category 3*, see [Section 2.1.4](#) and [Fig. 5](#)):

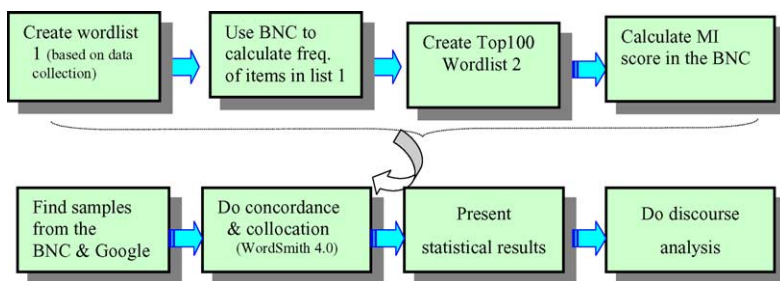


Fig. 3. Flowchart for the corpus and its data-processing.

### PITTER PATTER RAINDROPS



(revised from <http://www.hiraeth.com/youngembroiderers/raindrop.htm>, searched by Google)

Fig. 4. 'Pitter patter'.

### Hide and Seek Puppies: **Tic-Tac-Toe**

It's time to play **Tic-Tac-Toe!**



You are the puppy. The computer is the doghouse. Beat the computer at **Tic-Tac-Toe!** **Have fun!**

(<http://www.hideandseekpuppies.com/tictactoe/>)

(abbreviated text & picture for tic tac toe with the frequency of 123,000)

Fig. 5. 'Tic-tac-toe'.



The example “tic-tac-toe” integrates reliteration “tic-tac” into the triple alliteration “tic-tac-toe”. SS reduplication is used to highlight the loudness of sounds made by the keyboard and imaginative/metaphoric children’s laughter when they hit the triplet and win the game. Therefore ‘to play Tic-Tac-Toe’ is a signal to ‘*beat the computer at Tic-Tac-Toe and Have fun*’.

## 4. Results and sample analysis

### 4.1. The top 102 reduplication wordlist

The reduplication wordlist (1) with 673 tokens was first created for the pilot study. Then the frequency of occurrences for each token in the BNC was searched. Among those, 256 tokens did not occur in the BNC. Then the top 102 tokens were chosen to create wordlist 2. The results show that ‘either ... or ...’ is ranked the highest: 22,111 in all. ‘Ragbag’ (31) is ranked the lowest (Table 4).

#### 4.1.1. Findings: extra reduplicative forms

There are various forms of reduplication other than those of the basic three groups in Table 3. In addition to the previous sixteen subtypes of reduplication (cf. Appendix A), other categories can be found in the top 102, in the author’s corpus and elsewhere as shown below (also see Thun, 1963; Wescott, 1980; McCarthy, 1990; Moon, 1998; Carter, 1998):

- (1) [X conj. Y] (e.g., here and there; trick or treat; wine and dine);
- (2) [X<sub>1</sub> conj. X<sub>1</sub>] (e.g., on and on; so and so; such and such);
- (3) [X prep. Y] (e.g., Queen of Mean; tit for tat; top to toe);
- (4) [X<sub>1</sub> prep. X<sub>1</sub>] (e.g., arm in arm; an eye for an eye; face-to-face; face to face);
- (5) [X<sub>1</sub> in, X<sub>1</sub> out] (e.g., year in, year out; cf. in and out);
- (6) [X<sub>1</sub> X<sub>2</sub> X<sub>3</sub>]: triplets/trinomials (e.g., tic tac toe; Milly Molly Mandy);
- (7) [X-a-Y] (e.g., bric-a-brac; pit-a-pat; ting-a-ling).
- (8) [X the Y] (e.g., kick the bucket; off-the-cuff; under the counter).

A closer look at the above patterns reveals that they can be categorized into three broad groups (Table 5):

Multiple orthographic words show a higher frequency than hyphenated ones, except for samples of [X]-a-[Y].

### 4.2. Mutual information (MI) score

Mutual information (MI) is used for assessing collocational significance. The *MI score* can help us decide what to look for in a concordance. It compares the frequency of co-occurrence of two words in a given scope with their predicted frequency of co-occurrence. These words are randomly distributed in the corpus. The higher the MI score, the more genuine the association between two words (Oakes, 1998: 89). MI can be calculated using

Table 4

Wordlist 2 for top 102 items

Rank	Frequency	Reduplication	Rank	Frequency	Reduplication	Rank	Frequency	Reduplication
1	22,111	either Xi or Y (i = 0–25)	35	112	Rococo	69	45	birth and death
2	6,537	TV (/tivi/)	35	112	positive or negative	70	44	beck and call
3	4,614	neither Xi nor Y (i = 0–25)	37	110	facts and figures	71	43	mix and match
4	2,732	Hong Kong	38	102	no go	72	43	riff-raff
5	2,578	mummy	38	102	zig-zag	73	42	tip-top
6	1,205	in and out	40	97	hip hop	73	42	death and destruction
7	1,193	highlight	41	95	aye aye	75	41	stars and stripes
8	790	AC/DC (bisexual)	42	92	tit for tat	75	41	subject and object
9	708	here and there	43	89	hubbub	77	40	ta-ta
10	698	picnic	44	85	mother and daughter	78	39	Tartar
11	565	polo	45	78	zigzag	78	39	heaven and hell
12	556	mama	46	77	humdrum	78	39	chit-chat
13	514	cocoa	46	77	short shrift	81	38	father and grandfather
14	486	papa	46	77	sing-song	82	37	hard hat
15	480	once or twice	49	76	voodoo	82	37	wee wee
16	328	Fleet Street	50	74	plus or minus	84	36	hither and thither
17	266	hedgehog	51	73	bye-bye	84	36	hot pot
18	254	past and present	52	72	what not	86	35	fast and furious
19	248	first and foremost	53	69	willy-nilly	86	35	tit-for-tat
20	215	so and so	54	68	pots and pans	88	34	hurly-burly
21	212	this and that	55	64	Mayday	88	34	tutu
22	207	Royal Mail	56	58	bigger and better	90	33	brain drain
23	204	dada	57	55	criss-cross	90	33	cash and carry
24	195	top and bottom	58	54	high and dry	90	33	Marks and Sparks
25	156	upper and lower	58	54	Toto (pop group)	90	33	Pride and Prejudice
26	153	wear and tear	60	53	bric-a-brac	90	33	think tank
27	148	CV	60	53	lulu	90	33	top shop
28	140	chip shop	60	53	dodo	96	32	charivari
29	137	Coca Cola	63	51	rough and ready	96	32	now or never
30	135	grandstand	63	51	rough and tumble	98	31	Duran Duran
31	134	part and parcel	63	51	bits and bobs	98	31	fifty-fifty
32	132	oboe (“haut bois”)	66	50	fun run	98	31	King Kong
33	130	such and such	67	47	life and limb	98	31	mish-mash
34	121	tried and tested	68	45	Humpty Dumpty	98	31	ragbag

Table 5  
Lexical and phrasal structure of English reduplication

	Alternation	Examples
<b>(1) Full copy</b>		
$X_1$ {conj., prep} $X_1$	$[X_1]$ conj. $[X_1]$	on and on (503); so-and-so (115); so and so (207); such-and-such (67); such and such (129) all-in-all (7); all in all (320); arm-in-arm (29); arm in arm (63); an eye for an eye (10);
	$[X_1]$ prep. $[X_1]$	eye-to-eye (6); eye to eye (63); face-to-face (271, adj.); face to face (527, adv.); hand-in-hand (67); hand in hand (274); day-by-day (17); day by day (121); side-by-side (60); side by side (539); year after year (174)
	$[X_1]$ in, $[X_1]$ out	day in, day out (17); cf. in and out (1152); night in, night out (2); week in, week out (17); month in, month out (1); year in, year out (14)
<b>(2) Partial reduplication</b>		
$X$ {conj., prep, art} $Y$	$[X]$ conj. $[Y]$	here and there (673); this-and-that (2); this and that (202); trick-or-treat (1); trick or treat (15); wine and dine (10)
	$[X]$ prep. $[Y]$	Pride of Paddington (0); Queen of Mean (0); Taste of Texas (0) [names of restraint] tit-for-tat (35); tit for tat (92); top-to-toe (11); top to toe (24)
	$[X]$ -a- $[Y]$	bric-a-brac (52); bricabrac (1); pit-a-pat (1); pitapat (1); rat-a-tat (0); rat a tat (8); rope-a-dope (0) ting-a-ling (2); tingaling (1) ding-a-ling (3); dingaling (2); ding a ling (1)
	$[X]$ the $[Y]$	by the by (34); kick the bucket (7); off-the-cuff (18); off the cuff (17); under-the-counter (7); under the counter (35)
<b>(3) Triplet &amp; others</b>		
	$X_1$ $X_2$ $X_3$	tic-tac-toe (1); tic tac toe (3); Milly-Molly-Mandy (2); Milly Molly Mandy (2)

cf. <http://www.hradec.org/projd.html>; <http://www.trussel.com/flipflop.htm>.

Table 6  
MWU: MI score

No.	Binomial/ ablaut X + Y	(Frequency 1) $f(x)$	(Frequency 2) $f(y)$	(Joint frequency) $f(x, y)$	MI score	Group (high/low)
1	zig zag	36	24	19	21.1	High (MI score)
2	riff raff	81	21	6	18.4	
3	willy nilly	335	9	9	18.2	
4	<i>tit</i> for <i>tat</i> *	207	173	92	18.0	
5	mish mash	5	206	2	17.6	
6	hip hop	1,069	512	97	14.1	Low (MI score)
7	chit chat	60	1,269	6	12.9	
8	criss cross	12	7,382	7	12.9	
9	hum drum	323	957	3	9.9	
10	tip top	2,460	25,701	15	4.6	

the SARA tool or the following formula (Church and Hanks, 1990; Stubbs, 1995)<sup>29</sup>:

$$I = \frac{\log 2(f(x, y) \times N)}{f(x) \times f(y)}$$

- $I$  = MI score
- $f(x, y)$  is the collocation frequency
- $f(x)$  is the frequency of the first item (the query focus)
- $f(y)$  is the frequency of the collocate
- $N$  is the number of words in the corpus (corpus size;  $N = 100,106,008$  in the BNC)
  - If  $I(x; y) > 3$ , then the pairs tend to be ‘interesting’.
  - If  $I(x; y) \simeq 0$ , then the pairs are less interesting.
  - If  $I(x; y) \leq 0$ , then  $x$  and  $y$  are in complementary distribution.

Given Table 6, the MI score for MWU ablaut reduplication can be calculated using the provided formula. Basically, the higher the MI score, the more significant the association between two items. As mentioned earlier, the MI score is used for measuring collocational relations. If, on the basis of their MI scores, the above items are categorized into two groups, a higher group (1–5) and a lower one (6–10), a  $t$ -test shows a significant difference for both groups ( $p < .01$ ). The MI score also implies that the values of  $f(x)$  for the higher group tend to be bigger than those of  $f(y)$ , except the example ‘mish mash’. For the lower group, the values of  $f(x)$  tend to be smaller than those of  $f(y)$ , except ‘hip hop’.

MI provides a quick summary of what company words keep; it helps us decide what to look for in concordances (Church and Hanks, 1990). It is evident that while all the items of

<sup>29</sup> An MI score greater than 3 may indicate a significant collocational link (Church and Hanks, 1990). Thanks also go to Ming-wen Wu who calculated MI scores in Table 6 based on this formula and pointed out the calculation of MI scores using Sara is inaccurate.

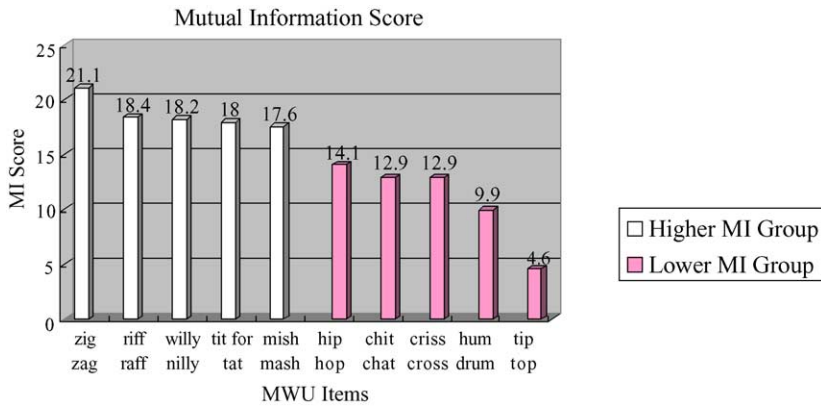


Fig. 6. Mutual information scores for reduplicative MWUs.

the higher groups (1–5) in Fig. 6 indicate higher collocation, a *t*-test points out that the lower part (6–10) has relatively lower probabilistic collocations ( $p < .01$ ). Two examples, ‘tit for tat’ ( $MI = 18.0$ ) and ‘chit chat’ ( $MI = 12.9$ ) are chosen to represent the higher and lower groups of *MI* scores in Fig. 6, respectively.

#### 4.3. Samples of discourse analysis reinforcing McCarthy (1988) and many others

##### 4.3.1. Integrating quantitative methods into qualitative analysis

The sample in Fig. 7 is an analysis of a text, extracted by Google search, with SS reduplication. Its lexical cohesion and word chain are analyzed (*Categories 4* and *5*, see Section 2.1.4)

The examples “chitchat” and “by guess and by gosh” are based on reliteration and alliteration at the lexical level and the phrasal level, respectively. The text is a manual for internet telephone installation. The title (‘No Idle Chitchat’) is closely tied up with its text. The topic exhibits strong lexical cohesion such as the Home Networking, Phone, HomeLine networks, HomePNA members, etc. Its advertising language is powerful and

**No Idle Chitchat: Home Networking on the Phone**

HomeLINE networks are based on the HomePNA 1.1 specification and should therefore interoperate with products from other HomePNA members like 3Com, Compaq, Lucent, IBM, and so on. That right there sounded pretty reassuring to me.

Read, Read, Read

Step one in the installation process: Read the manual! I've been doing this technology stuff long enough to know that the previous statement seems decidedly wimpy to me, but trust me on this: You don't want to tackle this product by guess and by gosh. Read the manual.

([http://www.techshopper.com/learn/networking\\_hardware/TSH20001107S0002](http://www.techshopper.com/learn/networking_hardware/TSH20001107S0002); based on Google search)  
(Note: frequency for No Idle Chitchat: 1,290; by guess and by gosh 94,900; texts abbreviated)

Fig. 7. Discourse analysis for SS reduplication.

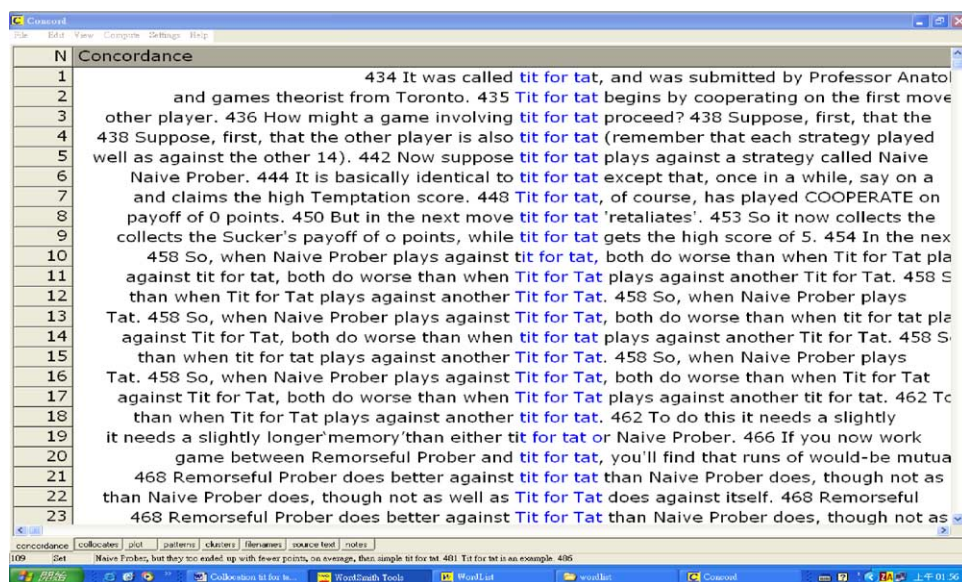


Fig. 8. The concordance results of *tit for tat* (using WordSmith 4.0).

contrastive with metaphoric sounds, e.g., ‘reassuring to me’ vs. ‘wimpy to me’; ‘Read, Read, Read’, ‘Read the manual!’, and ‘Read the manual!’ (cf. McCarthy, 1990: 55, 106; McCarthy, 1991: 64–117).

The next step is to extract examples or texts from the BNC and Google for the purpose of extended discourse analysis. For instance, the results of a search for ‘tit for tat’ (frequency of occurrence in the BNC: 92; Google: 91,800) are displayed in the ‘concordance’ results window using updated WordSmith 4.0 (cf. Scott, 2001 and later) (Fig. 8):

The extracted texts are constructed as a sub-corpus for ‘tit for tat’ with 2179 running tokens<sup>30</sup> in total; 2082 tokens were used for the wordlist. The average sentence length is 25.70 words. Table 7 provides a summary of the statistical information of ‘tit for tat’ in the extracted sub-corpus from the BNC.

The concordance lines and collocations of the sub-corpus are further analyzed in Table 8.

A closer look at concordance line 8 in Table 8 shows that the rightward collocates for [Tit for Tat] are ‘both do worse’ used to indicate ‘punishment given to someone in return for harm done to oneself.’ The immediate leftward collocate is ‘against’. The statistical results in Table 9 show the top tokens naturally ranked by *for/tit/tat*. ‘Tit for tat’ implies ‘in revenge’.

<sup>30</sup> The author found that different software packages would yield different statistical results.

Table 7

Summary of the statistical information for ‘tit for tat’

No.	Statistical items	No.	Statistical items
1	file size	11,678	9 mean word length (in characters)
2	tokens (running words) in text	2,179	10 word length std.dev.
3	tokens used for wordlist	2,082	11 sentences
4	types (distinct words)	588	12 mean (in words)
5	type/token ratio (TTR)	28.24	13 std.dev.
6	standardised TTR	33.90	14 paragraphs
7	standardised TTR std.dev.	46.74	15 mean (in words)
8	standardised TTR basis	1,000	

Note: The statistical results are calculated in terms of WordSmith 4.0.

Table 8

Concordance and Collocation for ‘tit for tat’ from the BNC

4. .... or Tat proceed? ARR 442 Now suppose	[[Tit for Tat]]	plays <u>against</u> a strategy called Naive
7. .... 458 So, when Naive Prober plays <u>against</u>	[[Tit for Tat]],	both do worse than when Tit for
8. .... st Tit for Tat, both do worse than when	[[Tit for Tat]]	plays <u>against</u> another Tit for Tat. ...
9. .... when Tit for Tat plays <u>against</u> another	[[Tit for Tat]].	ARR 462 To do this it needs a
11. .... 8 Remorseful Prober does better <u>against</u>	[[Tit for Tat]]	than Naive Prober does, though
12. .... aive Prober does, though not as well as	[[Tit for Tat]]	does <u>against</u> itself. ARR 481 Tit
13. .... or Tat does <u>against</u> itself. ARR 481	[[Tit for Tat]]	is an example. ARR 512 Programmers ...
28. .... Tat is not an ESS. ARR 595 Unlike	[[Tit for Tat]],	Always Cooperate is not stable <u>against</u> ...

Table 9

Wordlist for Top 30 tokens in the sub-corpus of ‘tit for tat’

Rank	Word	Frequency	%	Rank	Word	Frequency	%
1	FOR	114	5.23	16	NOT	15	0.69
2	TIT	100	4.59	17	OTHER	15	0.69
3	#	97	4.45	18	THIS	15	0.69
4	TAT	93	4.27	19	NICE	13	0.60
5	THE	91	4.18	20	AN	12	0.55
6	OF	64	2.94	21	BY	12	0.55
7	A	45	2.07	22	STRATEGIES	12	0.55
8	TO	43	1.97	23	STRATEGY	12	0.55
9	AND	38	1.74	24	BUT	11	0.50
10	IN	35	1.61	25	IF	11	0.50
11	IS	35	1.61	26	WITH	11	0.50
12	IT	30	1.38	27	<b>AGAINST</b>	<b>10</b>	<b>0.46</b>
13	THAT	21	0.96	28	ARE	10	0.46
14	AS	18	0.83	29	ESS	10	0.46
15	BE	16	0.73	30	ON	10	0.46

Note: The double cross, ‘#’, represents the code numbers, based on the analysis of WordSmith 4.0.

Table 10

Top 25 tokens for left and right collocation of 'tit for tat'

No.	Word	Total	Total Left	Total Right	L5	L4	L3	L2	L1	Center	R1	R2	R3	R4	R5
1	FOR	115	17	98	7	4	5	0	1	0	92	0	1	0	5
2	<b>TIT FOR TAT</b>	<b>92</b>	0	0	0	0	0	0	0	<b>92</b>	0	0	0	0	0
3	TAT	89	8	81	3	2	1	2	0	0	0	81	0	0	0
4	THE	24	19	5	4	5	6	1	3	0	0	0	2	2	1
5	A	21	16	5	3	5	5	0	3	0	0	0	0	4	1
6	IS	21	8	13	1	2	3	2	0	0	0	0	11	1	1
7	OF	20	15	5	3	2	0	4	6	0	0	0	1	2	2
8	TIT	14	7	7	2	5	0	0	0	0	0	0	0	5	2
9	TO	13	9	4	2	1	4	1	1	0	0	0	1	1	2
10	IN	12	7	5	2	3	1	1	0	0	0	0	0	2	3
11	AND	10	5	5	2	0	1	1	1	0	0	0	3	1	1
12	BE	10	6	4	2	2	1	1	0	0	0	0	0	4	0
13	THAT	10	7	3	0	2	0	1	4	0	0	0	0	2	1
14	THIS	10	6	4	1	2	0	1	2	0	0	0	0	1	3
15	AN	9	2	7	0	1	1	0	0	0	0	1	0	3	3
16	AS	9	7	2	1	0	3	0	3	0	0	0	2	0	0
17	NOT	9	4	5	0	2	1	1	0	0	0	0	0	4	1
18	SUSPICIOUS	9	6	3	0	0	0	0	6	0	0	0	2	1	0
19	BY	7	5	2	2	0	0	0	3	0	0	0	0	2	0
20	SO	7	4	3	1	0	1	2	0	0	0	0	0	1	2
21	<b>AGAINST</b>	<b>6</b>	<b>3</b>	<b>3</b>	0	0	0	<b>1</b>	<b>2</b>	0	0	0	0	<b>3</b>	0
22	COULD	6	3	3	0	3	0	0	0	0	0	0	3	0	0
23	DOES	5	1	4	0	0	1	0	0	0	0	0	4	0	0
24	INDIVIDUALS	5	0	5	0	0	0	0	0	0	0	0	5	0	0
25	LIKE	5	4	1	1	0	0	0	3	0	0	0	0	1	0

Note: The bidirectional collocations, five words for each, are based on the analysis of WordSmith 4.0.

When 'tit for tat' is analyzed both rightward and leftward, five words for each, lexically it tends to collocate with 'against' (10 times here), which demonstrates a semantic prosody of 'conflict'/'struggle' (see Table 10)<sup>31</sup>.

The advanced search is done for example sentences containing *against*. Both Table 10 and Fig. 9 display the distributions and concordance results of 'against' and 'tit for tat'. It is evident that 'against' can occur before or after 'tit for tat'. The remaining examples beyond the adjacent scope of five words can be traced as well in terms of this advanced search, as shown in Fig. 9.

One more relevant topic proposed by Moon (1998: 58) is to assess the significance and difference between observed frequency and expected frequency. First, the frequency of occurrences in the BNC for each token is depicted in Table 11.

The formula for calculating the expected frequency of 'tit for tat' is given in Fig. 10.

The expected frequency is calculated as shown in Fig. 11.

<sup>31</sup> Both MonoConc Pro and WordSmith 4.0 may generate slightly different statistical results for the same file when running their programming function, which will be discussed elsewhere.



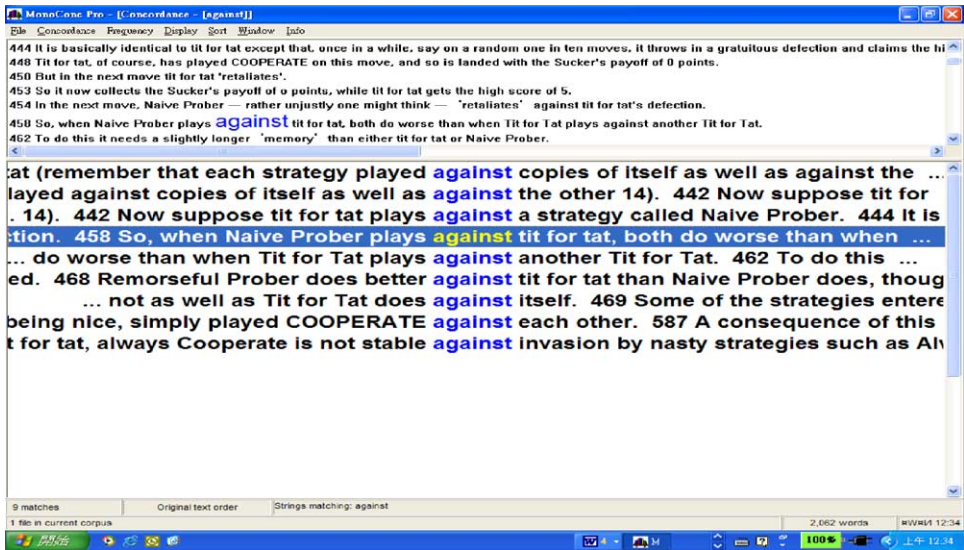


Fig. 9. The Concordance Results of ‘against’ in *tit for tat* (using MonoCron Pro).

Table 11  
Frequency of occurrences for ‘tit’ and ‘tat’ in all BNC and sub-corpus

Type/token	tit	tat	tit for tat	Total words in BNC
Frequency in all BNC	207	173	92	100,106,108
Frequency in sub-corpus	100	93	92	2,179 (running words) 2,082 (for wordlist)

Note: The results of the sub-corpus are based on the BNC search and the analysis of WordSmith 4.0.

$$E = \frac{\text{frequency word 1}}{\text{tokens in corpus}} \times \frac{\text{frequency word 2}}{\text{tokens in corpus}} \times \text{window}$$

\*\*Window’ is used for non-adjacent collocations.

Fig. 10. Formula for predicting frequency of non-adjacent collocations (Moon, 1998). Reproduced with written permission from Oxford University Press.

The expected value reveals that ‘tit for tat’ should appear 7 times for every 100 million words in the BNC; the actual observed frequency is 92 per 100 million words, as shown earlier. “Th[is] result gives the likelihood of events occurring at any given point in the corpus” (Moon, 1998: 58). Events here mean three-word sequences; the functional word ‘for’ is ignored when calculating non-adjacent collocations.

$$E = \frac{207}{100,106,008} \times \frac{173}{100,106,008} \times 2 = 7.15 \times 10^{-12}$$

Fig. 11. Calculating expected frequency of ‘tit for tat’.

## US, Russia in tit-for-tat 'cold war' expulsions

Martin Kettle in Washington and Amelia Gentleman in Moscow  
Friday March 23, 2001  
The Guardian

[Special report: Russia](#)  
[Special report: George Bush's America](#)

The **biggest barrage** of **tit-for-tat** diplomatic **expulsions** **between** Washington and Moscow since the end of the **cold war** was under way last night, sending US-Russian relations to a **new low** and providing the **Bush administration** with **its** first **big** test on the international stage.

<http://www.guardianunlimited.co.uk/russia/article/0,2763,461703,00.html>

frequency of 'tit for tat': 91,800 based on Google searching

Fig. 12. Reduplication and metaphoric excerpt of The Guardian (© The Guardian) reproduced with written permission from The Guardian.

#### 4.3.2. Samples for discourse analysis to repetition and reduplication

The next step is to do a discourse analysis using authentic materials extracted from websites. As seen in Fig. 12, the patterns of collocation, reduplication and repetition draw the readers' attention across the text. An analysis of the following text exhibits the features of polarity for conflicting parties, and other co-occurring sound patterns emerge:

This extract demonstrates how reduplication is used as an instrument of metaphor to enhance the hostile position of two *Powers*:

- 1) Topic: US vs. Russia; tit for tat; cold war (both countries, two counterparts)
- 2) Capital vs. Capital: Washington vs. Moscow (two capitals)
- 3) Special vs. Special (two reporters)
- 4) Special report: Russia vs. Special report: George Bush's America
- 5) Alliteration, reliteration and SS reduplication:

The topic for expressive vocabulary and the reiteration for **biggest barrage** of **tit-for-tat** are lexically cohesive and exhibit sound effects. Likewise, **tit-for-tat** diplomatic **expulsions**, **cold war** and **new low** indicate the polarity of two countries.

- 6) Country repetition (metonym): US, Washington, George Bush's America, Bush's Administration vs. Russia, Moscow; Russia, Moscow
- 7) Pronoun: its = [of] Bush's Administration
- 8) Sound repetition: B (bilabial plosive) sound repeated in 'biggest barrage', '**expulsion**'<sup>32</sup> and "the **Bush** administration with its first **big** test" to underscore the big trouble of Bush administration.

<sup>32</sup> An underlying unaspirated [p] sound follows /s/; it sounds like the devoiced bilabial /b/. It is well-known that English /p, t, k/ consonants become unaspirated after /s/ (the *English Unaspirated Rule*: /p, t, k/ → [-aspirated] / s \_\_\_\_).

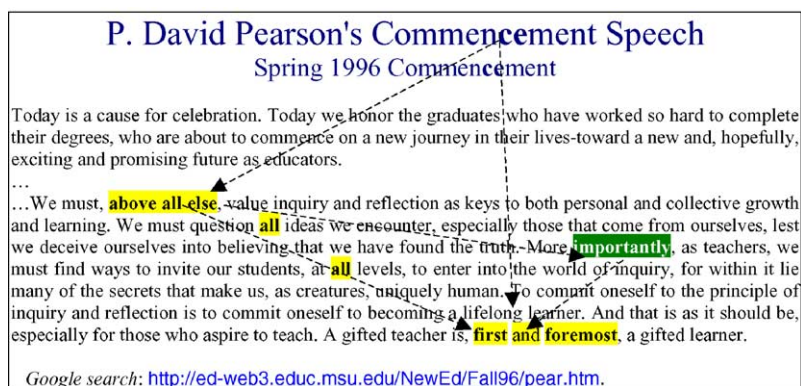


Fig. 13. Discourse analysis to repetition and reduplicative fixed expression. Used with written permission of the author.

One more feature of metaphoric expression is the pattern of bilabial onset-repetition **B...B...P** applied to underscore: *B-B-B...Bush is Poor*, suggesting how bad and poor Bush's situation was at that moment. Recall the slogan about Buick (see Section 2.3.3), "When Better Automobiles Are Built, Buick Will Build Them". There, the prominent attribute of the slogan was that it repeats the /b/-sound four times to call attention to its better quality. The pattern of bilabial onset-repetition **W...B** is used for emphasis: *We're Buick—We're Better*.

If we go further in terms of the data thrown up by the Google search engine and refer back to Table 3 and Appendix A, we will find that the message in Fig. 13 undergoes various types of reduplication at different levels.

An integrated approach using discourse analysis can also be applied to reduplicative fixed expressions. For instance, the fixed expression, *first and foremost* is given in Fig. 13 ( $f(n) = 120825$ ;  $f(c) = 604$ ;  $f(n, c) = 248$  in the BNC). The meaning of this freeze implies 'most importantly' or 'above all else', using a reduplicative *circumsonance* pattern, such as '*f...st*', in 'first and foremost' (Wang, 2001a, 2001b). Repetition functions as a kind of lexical cohesion, by stressing the close relationship between the overall topic and the individual lexemes. The lexical chain, 'commencement' → 'above all else' → 'more importantly' → 'first and foremost', displayed above, leads to emphasizing how the teaching and learning are important in this lecture (cf. McCarthy, 1990: 55, 106; McCarthy, 1991: 64–117). More details are given in the following:

#### 1) P. David Pearson's Commencement Speech

##### Spring 1996 Commencement

- headlines, including two special sounds, /s/, /p/ and combined /sp/
  - Pearson's Commencement Speech
- /p/-sound repetition → P. David Pearson's ... Speech...Spring...

- 2) Palindrome expressions at discourse level (Type A, see [Appendix A](#))
  - Contrastive headline-1 and headline-2 (Visual repetition)
  - mirror image-1: /p...p...s...s...s...sp/ → /sp...s/ (Mirror image)
  - mirror image-2: /s...sp/ → /sp...s/
 → Commencement Speech vs. Spring 1996 Commencement
- 3) Circumsonance at phrasal level (Types B and C)
  - /f...st/ sound pattern → ‘first and foremost’
- 4) Collocation at discourse level
  - Collocation: ‘above all else’ → ‘more importantly’ → ‘first and foremost’
  - Synonymous repetition
- 5) Lexical cohesion at discourse level (cf. [McCarthy, 1988](#))
  - Commencement implies the beginning of something, a graduation ceremony, thus very important as the lexical chain displays, ‘commencement’ → ‘above all else’ → ‘more importantly’ → ‘first and foremost’.
  - Synonymous repetition

The concurrent sound patterns show up repeatedly, especially the fricative /s/, implying “efficacy through the possession of abrasive qualities” ([Goddard, 1998](#): 83–84). The patterns and the onset/coda repetition of the /s/ sound are especially used in both headline and sub-headline. This emphasizes that commencement means something *new*; the graduates ‘are about to commence a *new* journey in their lives.’ It further implies that anything not related to the relationship between teaching and learning should be abandoned (or ‘abraded’), thus highlighting the significance of their relations and connections for education.

Both excerpts above exhibit a certain musical interplay between sound repetitions and various types of reiteration/synonyms. They demonstrate how R/R plays a striking role in these extracted texts ([Cook, 1989](#); [Tannen, 1989](#); [Carter, 1998](#)), thus confirming McCarthy’s arguments ([McCarthy, 1988](#); [McCarthy and Carter, 1994](#)).

## 5. Concluding remarks

The aim of this research, using corpus-based approaches, has been to explore SS reduplication in terms of its types, frequency, percentage coverage, concordance, collocation, and MI scores, and of its applications in the real world. The corpus with a ranked wordlist was constructed basically according to the framework of [McCarthy \(1990\)](#); additional discourse analysis models of lexical cohesion were employed to show that corpora can play an important role in the integration of formalism and functionalism. Different language levels and areas of use are discussed, such as language games, headlines, brand names, and advertisements.

This research attempts to shed new light on R/R analysis, especially at the discourse level, based on authentic samples. The widespread distribution of R/R can be found on, or extracted from websites and in daily occurrences elsewhere, not just in literary discourses. They are present functionally and pragmatically in all types of everyday language, not just in its specific or technical uses.

As to form, R/R is partly predictable and regular, corresponding to expected types of sound alternation, but exceptions remain. That is, the deterministic rules are incomplete. It is therefore important to recall that R/R is a probabilistic relationship, demonstrated by corpus probabilities (e.g., collocation). Statistics such as percentage coverage and frequency of occurrences in a corpus are required to reinforce the analysis and provide relevant arguments and research approaches. MI scores are useful references while choosing which MWU or SS reduplication to discuss. Further studies integrating probabilistic methods are definitely needed, in particular such that integrate corpus-based and other theories (Yip, 2000; Tao and McCarthy, 2001; Manning, 2003), so as to help us further explore R/R in terms of discoursal and pragmatic analysis.

## **Acknowledgements**

The author is deeply grateful to Michael McCarthy, Ronald Carter, Fengfu Tsao, John Ohala, Svenja Adolphs, Michael Hoey, Ellen Chen, Stuart Davis, Chinfa Lien, Norbert Schmitt, Vitas Raskevicius, Jacob Mey, Andie Warren, Stephen Kirton and three anonymous reviewers for their valuable comments, help or encouragement on earlier drafts of this paper. Thanks also go to Marianne Celce-Murcia, Diana Larsen-Freeman, Lesley Jeffries, Elite Olshtain, Rosamund Moon, David Pearson, Oxford University Press, The Guardian and Palgrave Macmillan for allowing the author to use their figures, tables or texts. This research is supported in part by a grant, NSC 90-2411-H-130-008, from the National Science Council.

## **Appendix A**

See [Table 12](#).

## **Appendix B**

See [Fig. 14](#).

## **Appendix C**

See [Fig. 15](#).

## **Appendix D. Integration of form, meaning and use (cf. DeCarrico and Larsen-Freeman, 2002)**

See [Fig. 16](#).

Table 12

Sixteen subtypes for the word-formation of English reduplication (cf. Wescott, 1980)

Types	Terms	Combination	Examples
<b>A</b>			
1	Archetypal palindrome	rightward and leftward readings are the same	kick, gig, gag
2-1/2-2	Deviation	anticipatory addition/ replacement	pip, cock, cackle, cuckold, agog, shush, mum, kook
2-3/2-4	palindrome	progressive addition/ replacement	
3	Further deviation: Quasi-palindrome	Phonotactical prohibition: e.g., nasal or liquid infix	gang, chick, konk, blab, barb, gawk, fop
<b>B</b>			
1	alliteration	onset-repetition	bamboo, gargoyle, gobbledy-gook
2	assonance	nucleus-repetition	dim-wit, flabbergasted,, speed-freak, loud-mouth
3	reliteration	coda- repetition	kickback, cock-sucker, muck-raker
<b>C</b>			
1	preliteration	alliteration + assonance	Sad Sack, Bugs Bunny
2	circumsonance	alliteration + reliteration	flimflam, wishy-washy, sing-song, group grope
3	rhyme	assonance + reliteration	rinky-dink, clap-trap, crum-bum, hocus-pocus
4	reduplication = total repetition	alliteration + assonance + reliteration	kaka, booboo, dum-dum
<b>D</b>			
1	palindromy and rhyme:	palindromy + assonance + reliteration	boob-tube
2	double palindromy and assonance:	2 palindromy + assonance	poppy-cock
3	palindromy, assonance, and rhyme	palindromy + assonance + reliteration	cock-sucking mother-fucking

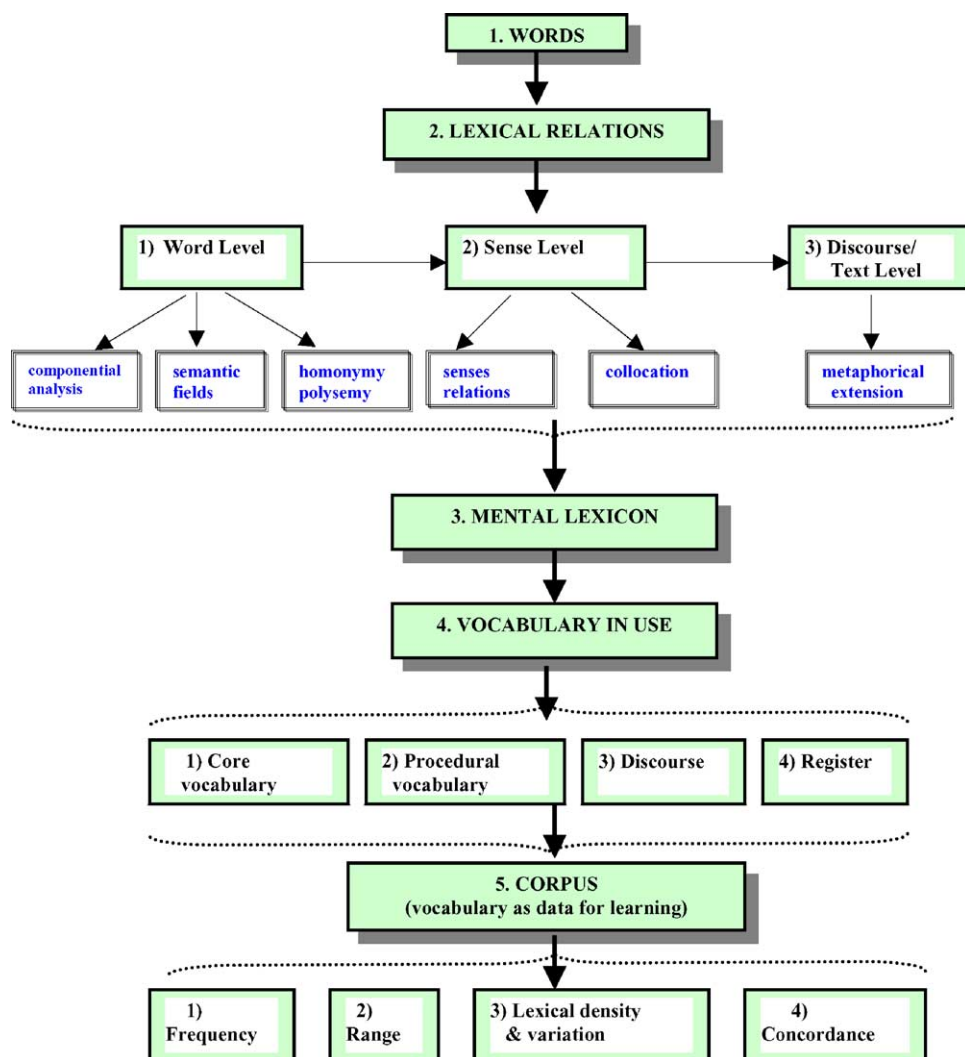


Fig. 14. McCarthy's framework (1990). Drawn by the author and revised by McCarthy.

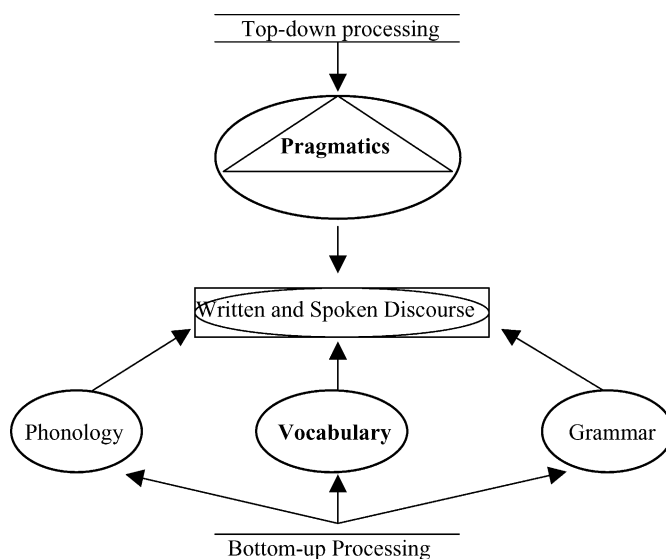


Fig. 15. Language knowledge framework (Celce-Murcia and Olshtain, 2000) *Note:* It is important for nonnative learners to make up for their lack of knowledge. They depend strongly on contextual features and prior knowledge to deal with new information, referred to as **top-down** or **knowledge-driven** explanation. **Bottom-up** or **data-driven** interpretation deals with the language processing of linguistic features such as spelling patterns, grammatical inflections, and word choices, and nonverbal cues such as gestures, illustrations, and so on. An effective learner or user can integrate both approaches into a suitable interpretation of what the speaker or writer expected. (Celce-Murcia and Olshtain, 2000: 13–14). Reproduced with written permission from the authors.

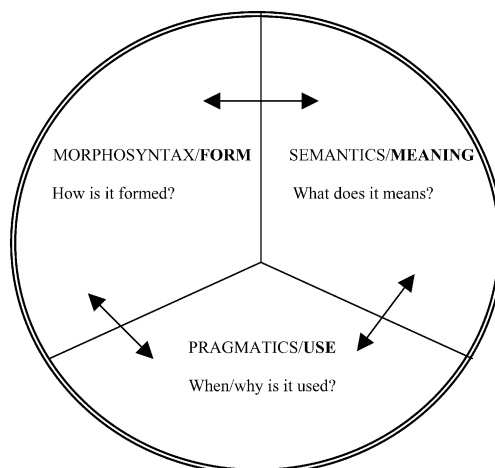


Fig. 16. Interconnected dimensions of grammar (Celce-Murcia and Larsen-Freeman, 1999). Reproduced with written permission from the authors.



## References

- Aston, Guy, Burnard, Lou, 1998. *The BNC Handbook*, Edinburgh University Press, Edinburgh.
- Beard, Adrian, 2000. *The Language of Politics*, Routledge, London.
- Biber, Douglas, Conrad, Susan, Reppen, Randi, 1998. *Corpus Linguistics: Investigating Structure and Use*, Cambridge University Press, Cambridge.
- Birdsong, David, 1995. Iconicity, markedness, and processing constraints in frozen locutions. In: Landsberg, M. (Ed.), *Syntactic Iconicity and Linguistic Freezes*, pp. 31–45.
- Bod, Rens, Hay, Jennifer, Jannedy, Stefanie, 2003. *Probabilistic Linguistics*, The MIT Press, Cambridge.
- Carter, Ronald, 1998. *Vocabulary: Applied Linguistic Perspectives*, Routledge, London.
- Carter, Ronald, McCarthy, Michael, 1999. The English get-passive in spoken discourse: description and implications for an interpersonal grammar. *English Language and Linguistics* 3 (1), 41–58.
- Carter, Ronald, McCarthy, Michael, 2004. Interactional language, creativity, and context. *Applied Linguistics* 25 (1), 62–88.
- Carter, Ronald, Goddard, Angela, Reah, Danuta, Sanger, Keith, Bowring, Maggie, 2001. *Working with Texts: A Core Introduction to Language Analysis*, second ed. Routledge, London.
- Celce-Murcia, Marianne, Larsen-Freeman, Diana, 1999. *The Grammar Book: An ESL/EFL Teacher's Course*, second ed. Heinle & Heinle, Boston.
- Celce-Murcia, Marianne, Olshtain, Elite, 2000. *Discourse and Context in Language Teaching*, Cambridge University Press, Cambridge.
- Church, Kenneth, Hanks, Patrick, 1990. Word association norms, mutual information, and lexicography. *Computational Linguistics* 16 (1), 22–29.
- Conrad, Susan, 1999. The importance of corpus-based research for language teachers. *System* 27, 1–18.
- Cook, Guy, 1989. *Discourse*, Oxford University Press, Oxford.
- Cook, Guy, 2000. *Language Play Language Learning*, Oxford University Press, Oxford.
- Cook, Guy, 2001. *The Discourse of Advertising*, Routledge, London.
- Cooper, William, Ross, John, 1975. World order. *Chicago Linguistic Society* 11, 63–111.
- Cutting, Joan, 2002. *Pragmatics and Discourse*, Routledge, London.
- Darnell, Michael, Moravcsik, Edith, Newmeyer, Frederick, Noonan, Michael, Wheatley, Kathleen (Eds.), 1998. *Functionalism and Formalism in Linguistics*, vol. i. John Benjamins Publishing Company, Amsterdam.
- Davis, Stuart, Hammond, Michael, 1995. On the status of onglides in American English. *Phonology* 12, 159–182.
- DeCarrico, Jeanette, Larsen-Freeman, Diane, 2002. Grammar. In: Schmitt, Norbert (Ed.), *An Introduction to Applied Linguistics*, Arnold, London, pp. 19–34.
- Fee, Jane, Ingram, David, 1982. Reduplication as a strategy of phonological development. *Journal of Child Language* 9, 41–54.
- Fenk-Oczlon, Gertraud, 1989. Word frequency and word order in freezes. *Linguistics* 27, 517–556.
- Ferguson, Charles A., 1983. Reduplication in child phonology. *Journal of Child Language* 10, 239–243.
- Finnegan, Ruth, 1977. *Oral Poetry: Its Nature, Significance, and Context*, Cambridge University Press, Cambridge.
- Garner, Bryan, 2002. *Elements of Legal Style*, Oxford University Press, Oxford.
- Goddard F Angela, 1998. *The Language of Advertising*, Routledge, London.
- Grant, Lynn, Bauer, Laurie, 2004. Criteria for re-defining idioms: are we barking up the wrong tree? *Applied Linguistics* 25 (1), 38–61.
- Green, Jonathon, 2000. *Cassell's Rhyming Slang*, Cassell & Co., London.
- Halliday, Michael, 1994. *An Introduction to Functional Grammar*, Edward Arnold, London.
- Hinton, Leanne, Nichols, Johanna, Ohala, John, 1994. *Sound Symbolism*, Cambridge University Press, Cambridge.
- Hoey, Michael, 1991. *Patterns of Lexis in Text*, Oxford University Press, Oxford.
- Ingram, David, 1974. Phonological rules in young children. *Journal of Child Language* 1, 49–64.
- Jeffries, Lesley, 1998. *Meaning in English*, Palgrave Macmillan, New York.
- Jespersen, Otto, 1933. Symbolic value of the vowel i. In: *Linguistica*, 1933. Also in *Selected Papers of Otto Jespersen in English, French and German*. Levin and Munksgaard, Copenhagen, pp. 283–303; or *Philologica*, 1, 1–19, 1922.

- Jespersen, Otto, 1964. *Language: Its Nature, Development and Origin*. The Norton Library, New York, or Allen and Unwin, London, 1922.
- Jakobson, Roman, Waugh, Linda, 1979. *The Sound Shape of Language*, Harvester Press, Sussex.
- Johnstone, Barbara, 1987. An introduction. Special issue on repetition. *TEXT* 7 (3), 205–214.
- Jurafsky, Daniel, 1996. Universal tendencies in the semantics of the diminutive. *Language* 72, 533–578.
- Károlyi, Krisztina, 2002. *Lexical Repetition in Text*, Peter Lang, Frankfurt am Main, Germany.
- Katamba, Francis, 1993. *Morphology*, Macmillan Press Ltd., London.
- Kennedy, Graeme, 1998. *An Introduction to Corpus Linguistics*, Addison-Wesley Longman Ltd., New York.
- Kimenyi, Alexandre, 1989. Reduplication and binomial expressions in English: a case of syntagmatic iconicity. In: Rauch, I., Carr, G. (Eds.), *The Semiotic Bridge: Trends from California*, Mouton de Gruyter, Berlin, pp. 347–353.
- Kirshenblatt-Gimblett, Barbara (Ed.), 1976. *Speech Play*. University of Pennsylvania Press, Philadelphia.
- Kadmon, Nirit, 2001. *Formal Pragmatics: Semantics, Pragmatics, Presupposition, and Focus*, Blackwell Publishers, Oxford.
- Lakoff, George, Johnson, Mark, 1980. *Metaphors we Live by*, University of Chicago Press, Chicago.
- Landsberg, Marge E. (Ed.), 1995a. *Syntactic Iconicity and Linguistic Freezes: The Human Dimension*. Mouton de Gruyter, New York.
- Landsberg, Marge E., 1995b. Semantic constraints on phonologically independent freezes. In: Landsberg (Ed.), *Syntactic Iconicity and Linguistic Freezes: The Human Dimension*. Mouton de Gruyter, New York, pp. 65–78.
- Lausberg, Heinrich, 1960/1998. *Handbook of the Literary Rhetoric: A Function for Literary Study*. *Handbuch der literarischen Rhetorik*. (Trans. by Mathew Bliss.) Brill, Leiden, Netherlands.
- Leech Geoffrey, N., 1966. *English in Advertising*, Longmans, Green & Co., London.
- Leech Geoffrey, N., 1983. *Principles of Pragmatics*, Longman Group, New York.
- Manning, Christopher D., 2003. Probabilistic syntax. In: Bod, Rens, Hay, Jennifer, Jannedy, Stefanie (Eds.), *Probabilistic Linguistics*. MIT Press, Cambridge, pp. 289–341.
- Marchand, Hans, 1969. *The Categories and Types of Present Day English Word-Formation: A Synchronic-Diachronic Approach*, Beck, Munich.
- Marantz, Alec, 1982. Re reduplication. *Linguistic Inquiry* 13, 435–482.
- McCarthy, Michael, 1988. Some vocabulary patterns in conversation. In: Carter, Ronald, McCarthy, Michael (Eds.), *Vocabulary and Language Teaching*, Longman, London, pp. 181–200.
- McCarthy, Michael, 1990. *Vocabulary*, Oxford University Press, Oxford.
- McCarthy, Michael, 1991. *Discourse Analysis for Language Teachers*, Cambridge University Press, Cambridge.
- McCarthy, Michael, 1998. *Spoken Language & Applied Linguistics*, Cambridge University Press, Cambridge.
- McCarthy, Michael, 2002. Good listenership made plain: British and American non-minimal response tokens in everyday conversation. In: Biber, D., Fitzpatrick, S., Reppen, R. (Eds.), *Using Corpora to Explore Linguistic Variation*, John Benjamins, Amsterdam, pp. 49–71.
- McCarthy, Michael, Carter, Ronald, 1994. *Language as Discourse: Perspectives for Language Teaching*, Longman Group Ltd., London.
- McCarthy, Michael, Walsh, Steve, 2003. *Discourse*. In: Nunan, D. (Ed.), *Classroom-Based Language Teaching Methodology*, McGraw-Hill, New York, pp. 173–195.
- McEnery, Tony, Wilson, Andrew, 2001. *Corpus Linguistics*, second ed. Edinburgh University Press, Edinburgh.
- McRae, John, 1998. *The Language of Poetry*, Routledge, London.
- Minkova, Donka, 2000. Reduplication and syllable well-formedness: the criss-crossing of phonology and semantics. In: Eleventh International Conference on English Historical Linguistics (<http://www.uvigo.es/webs/h04/jperez/icehl/phono.htm#minkova>).
- Moon, Rosamund, 1997. Vocabulary connections: multi-word items in English. In: Schmitt, N., McCarthy, M. (Eds.), *Vocabulary*, Cambridge University Press, Cambridge, pp. 40–63.
- Moon, Rosamund, 1998. *Fixed Expressions and Idioms in English*, Oxford University Press, Oxford.
- Moravcsik, Edith, Wheatly, Kathleen, 1998. Introduction. In: Darnell, M., Moravcsik, E., Newmeyer, F., Noonan, M., Wheatley, K. (Eds.), *Functionalism and Formalism in Linguistics*, vol. i. John Benjamins Publishing Company, Amsterdam, pp. 1–7.
- Nation, Paul, 1982. Beginning to learn foreign vocabulary: a review of the research. *RELJ* 13 (1), 14–36.

- Nation, Paul, Meara, Paul, 2002. Vocabulary. In: Schmitt, N. (Ed.), *An Introduction to Applied Linguistics*, Arnold, London, pp. 35–54.
- Oakes, Michael, 1998. *Statistics for Corpus Linguistics*, Edinburgh University Press, Edinburgh.
- Ohala, John, 2001. Sound symbolism: an explanatory theory from ethnology. (ms) A lecture in 2001 LSA Summer Institute at UCSB. In: *Proceedings of Seoul International Conference on Linguistics*. Linguistic Society of Korea, Seoul, pp. 98–103.
- O’Keeffe, Anne, Farr, Fiona, 2003. Using language corpora in initial teacher education: pedagogic issues and practical applications. *TESOL Quarterly* 37 (3), 389–418.
- Persson, Gunnar, 1974. *Repetition in English*. Ph.D. dissertation, Uppsala University.
- Pinker, Steven, Birdsong, David, 1979. Speaker’s sensitivity to rules of frozen word order. *Journal of Verbal Learning and Verbal Behavior* 18, 497–508.
- Pomorska, Krystyna, Chodakowska, Elzbieta, McLean, Hugh, Vine, Brent, 1987. *Language, Poetry and Poetics*, Mouton de Gruyter, Berlin.
- Regier, Terry, 1998. Reduplication and the arbitrariness of the sign. In: *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society*, pp. 887–892.
- Rundell, Michael, 1995. About language. *Modern English Teacher* 4 (2), 13–15.
- Sanches, Mary, Kirshenblatt-Gimblett, Barbara, 1976. Children’s traditional speech play and child language. In: Kirshenblatt-Gimblett (Ed.), *Speech Play*, pp. 65–100.
- Sapir, Edward, 1929. A study of phonetic symbolism. *Journal of Experimental Psychology* 12, 225–239.
- Schmitt, Norbert, 2000. *Vocabulary in Language Teaching*, Cambridge University Press, Cambridge.
- Schwartz, Richard, Leonard, Laurence, Wilcox, Jeanne, Folger, Karen, 1980. Again and again: reduplication in child phonology. *Journal of Child Language* 7, 75–87.
- Scott, Mike, et al., 2001. Comparing corpora and identifying key words, collocations, frequency distributions through the wordsmith tools suite of computer programs. In: Ghadessy, Mohsen (Ed.), *Small Corpus Studies and ELT*, John Benjamins Publishing Company, Amsterdam, pp. 47–67.
- Short, Mick, 1996. *Exploring the Language of Poems Plays and Prose*, Longman, London.
- Sinclair, John, 1991. *Corpus, Concordance, Collocation*, Oxford University Press, Oxford.
- Sinclair, John, et al., 1997. *Wichmann, Anne (Ed.), Corpus Evidence in Language Description*, Addison Wesley Longman, London, pp. 27–39.
- Spöttl, Carol, McCarthy, Michael, et al., 2003. Formulaic utterances in the multilingual context. In: Cenoz, Jasone (Ed.), *The Multilingual Lexicon*, Kluwer Academic Publishers, Dordrecht, pp. 133–151.
- Steriade, Donca, 1988. Reduplication and syllable transfer in Sanskrit and elsewhere. *Phonology* 5, 73–155.
- Stubbs, Michael, 1995. Collocations and semantic profiles: on the cause of the trouble with quantitative studies. *Functions of Language* 2 (1), 23–55.
- Tao, Hongyin, McCarthy, Michael, 2001. Understanding non-restrictive *which*-clauses in spoken English, which is not an easy thing. *Language Sciences* 23, 651–677.
- Tannen, Deborah, 1987. Repetition in conversation as spontaneous formulaicity. *TEXT* 7 (3), 215–243.
- Tannen, Deborah, 1989. *Talking Voices*, Cambridge University Press, Cambridge.
- Thun, Nils, 1963. *Reduplicative Words in English*, Carl Bloms Boktryckeri, Lund.
- Wang, Shih-ping, 2001a. Integrating corpus-based and vocabulary learning approaches into a linguistics project. Paper presented at the 46th International Linguistics Association Annual Meeting, New York University, New York.
- Wang, Shih-ping, 2001b. Reduplication and repetition in applied linguistics. In: *Proceedings of the 2001 International Conference on the Application of English Teaching*. Crane Publishing Co., Taipei, Taiwan, pp. 200–222.
- Wang, Shih-ping, 2003a. Mutual information and corpus-based approaches to reduplicative fixed expressions. *Corpus Linguistics* 2003 (2), 869–878, University of Lancaster, UK.
- Wang, Shih-ping, 2003b. Corpus-based approaches to frequency and word association for English vocabulary and multiword units. Paper presented at the 36th Annual BAAL Meeting, Leeds University, 4–6 September 2003.
- Waugh, Linda, et al., 1987. On the sound and shape of language: mediacy and immediacy. In: Pomorska, K. (Ed.), *Language, Poetry and Poetics*, Mouton de Gruyter, Berlin.
- Wescott, Roger, 1980. *Sound and Sense: Linguistic Essays on Phonosemic Subjects*, Jupiter Press, Lake Bluff, IL.

- Wales, Hugh, Gentry, Dwight, Wales, Max (Eds.), 1958. *Advertising Copy, Layout and Typography*. Ronald Press, New York.
- Widdowson, H.G., 1989. Knowledge of language and ability for use. *Applied Linguistics* 10, 128–137.
- Williams, Christopher, 2004. Legal English, the modal revolution and ELT. Manual script and handout for a talk at Nottingham University, 11 February 2004.
- Wray, Alison, 2000. Formulaic sequences in second language teaching: principle and practice. *Applied Linguistics* 21 (4), 463–489.
- Wray, Alison, 2002. *Formulaic Language and the Lexicon*, Cambridge University Press, Cambridge.
- Yip, Moira, 2000. Reduplication as alliteration and rhyme. *GLOT International* 4 (8), 1–7.

**Shih-ping Wang** is a Senior Lecture in the Department of Applied English, Ming Chuan University, Taoyuan, Taiwan. He is also a PhD candidate in Applied Linguistics, School of English Studies at University of Nottingham, UK, working on corpora, fixed expressions and discourse analysis with his supervisor, Professor Michael McCarthy. His research interests include corpus-based approaches to language studies, discourse analysis, fixed expressions, vocabulary and sound symbolism. He is currently the recipient of *Dissertation Fellowship of CCKF for International Scholarly Exchange* (2004–2005; DF025-U-03). From September 2005, his mailing address will be: Department of Applied English, Ming Chuan University, #5 De-ming Road, Gui-shan, Taoyuan County, Taiwan.