<table>
<thead>
<tr>
<th>Title</th>
<th>Collocational aspects of near-synonyms: Illustrations from a small corpus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author(s)</td>
<td>Backhouse, Anthony E.</td>
</tr>
<tr>
<td>Citation</td>
<td>北海道大学留学生センター紀要 = Journal of International Student Center, Hokkaido University, 7: 50-68</td>
</tr>
<tr>
<td>Issue Date</td>
<td>2003-12</td>
</tr>
<tr>
<td>Doc URL</td>
<td><a href="http://hdl.handle.net/2115/45637">http://hdl.handle.net/2115/45637</a></td>
</tr>
<tr>
<td>Type</td>
<td>bulletin (article)</td>
</tr>
<tr>
<td>File Information</td>
<td>BISC007_004.pdf</td>
</tr>
</tbody>
</table>

Hokkaido University Collection of Scholarly and Academic Papers: HUSCAP
Collocational aspects of near-synonyms: Illustrations from a small corpus

Anthony E Backhouse

Introduction
A central, though neglected, goal of language learning is the acquisition
of lexical semantic and pragmatic competence – the ability to use and understand lexical items in the same way as native speakers. Learners’ dictionaries, both monolingual and bilingual, have a major role to play in aiding this process, and as a basic requirement they need to provide language learners with a full range of information required in approaching this task.

A familiar aspect of the difficulties faced by learners in this area is the fact that languages are full of ‘near-synonyms’, i.e. lexical items whose meaning and use overlap in varying degrees, and distinguishing clearly between such items is an important task for learners’ lexicography. In many cases, items of this kind may be used to refer to similar phenomena and share translation equivalents in other languages, but differ in wider aspects of meaning and use. Such cases illustrate particularly clearly the fact that full lexical competence requires control of a variety of factors involved in the natural use of vocabulary, and that lexical descriptions for learners must accordingly go beyond simple definitions and glosses. Native speakers gain control of the lexicon on the basis of familiarity with lexical items used in real contexts, and significant aspects of such contexts need to be drawn upon in treatments in learners’ dictionaries.

By way of illustration here we take up two groups of near-synonyms in Japanese, and focus upon their collocational properties, based on attested examples drawn from a small corpus study. Under ‘collocation’ we include, in particular, co-occurrence with other lexical items in syntactic constructions, as well as broader significant linguistic aspects of contexts in which lexical items are naturally used. Our purpose is to show that studies of collocation can reveal a range of information relevant to natural lexical use, and that the provision of such information in learners’ dictionaries can enhance their usefulness as guides to lexical competence. Cross-linguistic focus is on English, as the author’s native language.

The two groups are (1) nezumi, haiiro, gure; and (2) shoseki, tosho, shomotsu (cf common English translation equivalents grey (1), book (2)). Attested examples for each item were obtained from a lexical search of internet homepage entries conducted via http://search.yahoo.co.jp/. The
first 100 entries for each item were collected, from which examples adjudged as unsuitable (cases of homography, names, repeated entries, examples where terms were mentioned rather than used, and examples whose reference was deemed to be unclear) were excluded to form the final corpus. As a lexicographical baseline, the treatment of these items in a representative monolingual (non-learners') Japanese dictionary (Kindaichi et al 1997) is also referred to.

1 Nezumiira, haiiro, guree

These three items all translate as English grey in most contexts. (In the case of nezumiira, morphologically nezumi 'rat/mouse' + iro 'colour', this is a cause of surprise for (some) English speakers, for whom mice are prototypically brown; cf mousy/mousey, applied to the colour of hair: 'Mousy hair is a dull light brown colour' (Cobuild 2001).) In a standard dictionary treatment, Kindaichi et al 1997 characterizes both haiiro and nezumiira as pale black colours, with haiiro being close to white and nezumiira showing a blue-white tinge (aojiromi); cigarette ash and concrete are given as typical examples for haiiro, and ienezumi for nezumiira. Haiiro is described as also occurring in various extended unfavourable meanings, as in the example haiiro no jinsei. Guree is defined as 'haiiro, nezumiira'. Note in particular that while this treatment indicates that guree overlaps in denotation with haiiro and nezumiira, it leaves open the general question of the contexts in which the respective items are naturally used.

Different types of lexical items focus attention on different aspects of collocations. For colour terms, which primarily describe qualities of entities, important collocational information includes the range of nouns with which these terms characteristically co-occur as modifiers or predicates. Noun collocations of these types provide us with an indication of what range of 'things' are naturally described as being haiiro, nezumiira or guree; note that on the basis of the dictionary description above, in the case of haiiro we may expect that these will include abstract 'things'. In addition, lexical items from a single domain of meaning (colour terms, in this case) may co-occur in language use, and the patterns of co-occurrence
found are also of interest: Do haiiro, nezumiiro and guree themselves co-occur, and if so in what patterns? Do the three terms differ in terms of their patterns of co-occurrence with other colour terms?

1.1 Guree

The corpus yielded 41 separate references for guree and composite items (such as gureekei, chakooruguree etc). (Since our concern is with collocational range, collocation with a given lexical item, whether single or repeated, in one entry is counted as one reference.) The nouns involved, arranged in rough semantic groups, are set out below, with numbers in parentheses indicating numbers of references.

Clothing (14)

*tiishatsu* (2)
*paakaa* (2)
*besuto*
*naiki* (shoes)
*nekutai*
*kooto*
*serafimu* (women’s suit)
*rongunikka* (trousers)
*pantsu*
*hatto*
*waishatsu*
*hankachi*

Computer-related (8)

*mausu* (3)
*konpyuntaa*
*geemukontoroora*
*tenpureeto*
*PCshuuenkiki*
*hyooji*
Personal accessories (6)
*baggu* (3)
*ferutokoroten* (bag)
*orisaiifu*
*bureeku* (backpack)

Household goods (3)
*maajantaku*
*reizooko*
*herusumeeta*

Paper products (3)
*funtoo*
*memo*
*meishiyooshi*

Stone types (3)
*mikageishi* (2)
*jari*

Personal decoration (2)
*paaru*
*neiruchippu*

Others (2)
*shiraga*
*kage*

The overwhelming use of *guree* in the corpus (36/41 references) is to specify the colour of manufactured objects (here, being advertised for sale), among which items of personal apparel (clothing, accessories, decorations), computers, and household goods and appliances predominate. Note that the
items of apparel all belong to the modern, ‘Western’ sector (cf Miura 1979: 62-3). Cases where *guree* is applied to natural colours are restricted to types of stone (being marketed as decorative products). Under Others, in the example of *shiraga* the discussion concerns dyed colours, and the shadows (*kage*) are those in photographs; both these examples involve co-occurrence with other ‘grey’ terms, as described below.

There are four cases in the corpus where *guree* co-occurs with other ‘grey’ terms. An entry introducing a photographic exhibition as *Haiiro to kuro no arenjimento* contains the later phrase *kuro ka guree no kage* apparently referring to the same features; we may conclude from this example that, in some cases at least, the referential ranges of *haiiro* and *guree* overlap. Elsewhere, *ferutokoTOten* (a bag) is followed by the descriptor *nezumiiiro*, which is in turn followed by the description *Koi guree desu*. The *shiraga* entry contains the sentence *Shiraga o makkuro ni somem no de wa naku guree (nezumiiro) ni suru no desu*. Finally, an entry for *baggu* contains the description *guree (nezumiiro)*. Together these three examples suggest clearly that *nezumiiiro* denotes a narrower range of colour within *guree*, namely darker shades of grey.

The complete list of colour terms which co-occur with *guree* is as follows: *nezumiiiro* (3), *haiiro*, *chakooruguree*, *chakooru*, *raitoshirubaa*; *burakkku* (3), *kuro* (2), *makkuro*; *howaito* (3), *shiro* (2), *afu*; *buruu* (3), *kon* (2), *ao*, *neibii*; *kiirokei*, *kikei*, *ieeroo*; *beeeju* (2), *raitobeeju*, *kogecha*; *murasakikei*, *paapuru*; *aka*; *pinku*; *orenji*; *midorikei*, *guriiin*, *aasuguriin*. A striking feature is the strong representation of foreign loans across the spectrum; notably, these outnumber even basic native colour terms in the case of *burakkku*, *howaito* and *buruu* (although *aka* is an exception in not being ousted by *reddo*).

1.2 *Nezumiiiro*

69 references occurred in the corpus. Noun collocations are indicated below.

Clothing (20)
(seifuku no) kiji (2)
obi (2)
chokki (2)
fuku
seifuku
kimono
(kimono no) gara
somemono
koppori
habutae
hoomongi
fundoshi
katakake
booshi
daffurukooto
suutsu
tiishatsu

Paint etc (10)
supuree (2)
puraimaa (2)
hausupeinto
tosoo
toryoo
(kantei no) toshoku
kureyon
jinbutsuga

Animals (4)
nezumi
chinchira
sorasuzume no ohire (changes from yellow)
(tora no) sutoraipu
Natural objects (3)

ishii (3)

Natural phenomena (5)

sora (2)

ame (2)

kami (due to light)

Artefacts (22)

fuusen (in people’s hearts)

muigurumi

hankachi (dirty)

yane

talemono

mooru (moulding)

jiipu

tooki (lamp)

hasa

gohyakuuenkooka

kaishubako

baggu

ferutokoroten (bag)

kiithorudaa

shiru

kamipakku

atsugami

itagami

(konpyuuta no) keeburu

(konpyuntagamen no) sen

mausu

roorodekkusu

Others (5)
As with *guree*, the majority of noun collocates (52/69) denote manufactured objects (clothing, paint, and other artefacts), but their range is wider: traditional clothing is well represented, and uniforms of various kinds are evidently commonly described as *nezumiiro; nezumiiro* also appears to be in common use with reference to paint; and the range of other artefacts extends beyond personal and household goods to cases such as *yane* and *tatemono*. In addition, unlike *guree*, *nezumiiro* is readily applied to natural colours, as with animals and the effects of natural light: stones appear among the collocates of both terms, but whereas the examples under *guree* refer to (processed?) types of stone being marketed for decorative purposes, the references here are to the colour of stones found in their natural state. Under Others, *echiopia no shushoku* is perhaps a further example of a natural colour; *nippon* and *kuni* involve figurative uses; *kao* relates to a dark-skinned person who has applied some white substance to his/her face; and the *shiraga* example involves dyeing, as mentioned under *guree* above. Overall, *nezumiiro* is thus applied to a wider range of entities than *guree* in the corpus, and takes in natural as well as manufactured colours.

Of five examples in the corpus where *nezumiiro* co-occurs with other 'grey' terms, three involve *guree* and were commented on in 1.1. One co-occurrence with *haiiro* relates to the stripes of a photographed tiger, described as *nezumiiro to haiiro no sutoraipu;* the second describes the effects of gradually darkening light on a sheet of writing paper: *Hi wa zunzun kurete iku no datta. Haiiro kara nezumiiro ni, nezumiiro kara sumiiro ni ... hiru no hikari wa yoru no yami ni kawaroo to shite ita.* Together the two examples show *nezumiiro* and *haiiro* functioning as contrast terms, with *haiiro* denoting a lighter, and *nezumiiro* a darker shade of colour.
The following colour terms co-occur with *nezumiiro* in the corpus: *guree* (3), *haiiro* (2), *usunezumi*, *chakooruguree*, *gin’iro*, *zooiro*; *kuro* (5), *kokushoku* (3), *kuroi* (2), *makkuro*, *sumiiro*; *shiro* (4), *shiroi* (4), *hakushoku* (2), *kurimuiiro*; *ao*, *aoro*, *bunuiiro*, *ai*, *aikeiilo* (no *iro*), *mizuiro*, *nookonshoku*, *kiiro* (4), *kiiroki*, *ki’iro*, *karashiirio*, *karashi*; *chairo* (6), *kilsuneiro*, *sabiirio*, *akasabiirio*, *usucha*, *kogecha*, *beeu*, *chokoreetoiro*; *aka* (2), *akai*, *enji*, *kuroenji*, *entanshoku*; *pinku*, *baraio*, *usumomoiro*; *murasaki*, *usumurasaki*, *usuusumurasaki*, *usuakamurasaki*; *orenji* (3), *daidaiiro*; *midori* (2), *kimidori* (2), *midoriirio*, *guriin*, *usukokeiro*. The most notable contrast with *guree* is the reduced prevalence of foreign loans with, in particular, neither *burakkku* nor *howailo* occurring here: whereas *guree* tends to co-occur with *burakkku*/*kuro* and *howailo*/*shiro*, *nezumiiro* favours *kuro*/*kokushoku*/*kuroi* and *shiro*/*shiroi*/*hakushoku*.

1.3 *Haiiro*

The corpus yielded 21 clear references for *haiiro*. Noun collocations are shown below.

Natural colours (12)

- *kabi* (2)
- *shokubutsu*
- *(tora no) sutoraipu*
- *noomiso*
- *noosaiboo*
- *kage* (as photographed)
- *kami* (due to light)
- *tsuchi*
- *yuki* (due to *jin’en*)
- *machi* (streets, due to *jin’en*)
- *ichinichi* (due to *kemuri*)

Artificial colours (2)

- *(tekisuto to aikon no) hyooji*
Figurative uses (7)
sora (2)
umi
machi (town)
sekai
nikki
watashi

While the number of clear examples is insufficient to support strong conclusions, they suggest that haiiro is applied predominantly to naturally occurring colours, less commonly to artificially produced colours (although for some examples there may be room for argument on the choice of placement within these categories). In addition, as indicated in Kindaichi et al, figurative uses account for a relatively high proportion (7/21) of references of haiiro: this is clear in the case of nouns which denote abstract entities like sekai, nikki and watashi, but even in concrete cases here the context indicates that it is the joyless mood of the viewer that is centrally involved, rather than necessarily the literal colour of the referents.

The three cases of co-occurrence of 'grey' terms involving haiiro (one with guree, and two with nezumiiro) have been described under 1.1 and 1.2. Co-occurrences of all colour terms are as follows: nezumiiro (2), guree; kuro (2), sumiiro; shiro; aka; ao, buruu. As with nezumiiro, burakku and howaito are absent.

1.4 Summary

Occurrences in the corpus have provided an indication of the range of nouns which these terms are used to describe. The clearest result is that guree is used predominantly to describe the colour of manufactured, 'modern' articles. Nezumiiro is likewise widely used with respect to manufactured articles, including traditional Japanese clothing, but also
describes natural colours, while haiiro appears rather to favour natural
colours as well as figurative uses. Co-occurrence data has also thrown light
on aspects of denotation: both haiiro and nezumiiro fall within the range of
guree, and there is clear evidence that the two terms may function in
contrast, denoting lighter and darker shades respectively.

2 Shoseki, tosho, shomotsu

These three items may be characterized as 'difficult' counterparts of hon,
all of which generally translate as book in English. By 'difficult' we refer to
the fact that, unlike hon which occurs commonly in a wide range of genres
including everyday conversation, these items are characteristic of more
formal, written-language styles. There is a widespread trend in the
Japanese lexicon for an everyday item (commonly, although not in this case,
from the native sector of the vocabulary) to have formal counterparts
(commonly from the Sino-Japanese sector) which also possess a narrower
semantic range (cf Miyajima 1988). While shoseki, tosho and shomotsu are
undoubtedly 'near-synonyms', we assume once again that they differ in
their conditions for natural use, and our main purpose here is to survey
collocational evidence for this. We focus in particular on three main aspects
of their behaviour. Firstly we consider the range of predicates with which
these nouns collocate as complements: what processes do they typically
participate in, as subjects, objects etc? Secondly we look at contexts in
which they co-occur as coordinate terms with other nouns: of what sets of
things do they commonly appear as members? Finally we take up aspects of
their compositional behaviour, mainly in compounds.

In lexicography, Kindaichi et al 1997 defines shoseki as hon, viewed as a
source of personal knowledge and as an object enriching everyday life.
Tosho is given two sub-entries: firstly, as a 'kangoteki' (i.e. stylistically
formal, Sino-Japanese) expression with the same meaning as shomotsu; and
secondly, as a collective term for books and other documentary resources
held in public facilities. Shomotsu is defined as a 'kangoteki' expression
with the same meaning as hon. Aside from the stylistic properties implied
by 'kangoteki' for shomotsu and for one use of tosho (although not for
shoseki, as in our characterization here), this treatment suggests that shoseki centres on the edificational aspect of books, tosho (in one of its uses) on their role as a public resource, while no specific semantic focus is indicated for shomotsu.

2.1 Shoseki

The corpus contained 105 references of shoseki (including composites). 25 references involved collocations of the noun shoseki as subject or object argument with predicates (verbs, verbal nouns + suru, or verbal nouns alone (e.g. shoseki no kensaku)). These are listed below, arranged by rough semantic groupings of predicates.

Selling (7)
- hatsubai (2)
- hatsubai suru
- hanbai suru (2)
- tsuuhan
- purezento

Buying (5)
- koonyuu (2)
- gokoonyuu
- omolome
- gokiboo

Information provision (5)
- shookai (2)
- goannai
- shuuroku suru
- matomeru

Information search (4)
- kensaku (2)
A majority of the references of shoseki in these collocations are to books as commercial products: shoseki are bought, sold, delivered, stocked, and compiled (in this instance, by the editorial department of a book publisher). Shoseki are also the target of information provision and search; much of this, although not all, is provided by publishers and booksellers and sought by prospective customers.

18 references of shoseki, either as an independent word or in composites, occurred in co-ordinate contexts involving sets of nouns. The commonest partner terms were zasshi (7), DVD (5), CD (4), indicating that shoseki often figures as one member among a range of commercially available sources of information and entertainment.

Finally we consider the behaviour of shoseki in composites. In general, caution is needed in comparing the meaning and use of words as free items directly with their behaviour as elements in compounds and derivatives, since word-formation commonly involves additional restrictions of various kinds (for example, on the combination of items from different etymological sectors of the vocabulary) which are not relevant above the level of the word. At the same time, however, consideration of composites of shoseki, tosho and shomotsu is clearly a potential source of additional comparative information of interest for these three terms. Shoseki appears in a large
number of composite items, among which we focus on compounds with *shoseki* as the first or second element, exemplified respectively by *shosekikoonyuu* and *ippanshoseki*. The majority of *shoseki* compounds show a complement-predicate structure, and occurring examples echo the collocations found above: the commonest compound of this type is *shosekikensaku*, followed by *shosekiannai* and *shosekishookai*. *-shoseki* compounds are generally of modifier–head structure, and overwhelmingly the commonest example in the corpus is *denshishoseki*, followed by *shinkanshoseki* and *kanrenshoseki*: compounds indicating subject matter interestingly appear to favour alphabetisms as the first element: *ITshoseki, UFOshoseki, SFshoseki*.

2.2 *Tosho*

Of the 71 references of *tosho* and composites in the corpus, 11 involved the noun *tosho* functioning as subject or object argument of predicates, listed below.

Information provision (3)

*keisai suru*

*shuuuroku*

*shookai*

Information search (3)

*kensaku suru*

*mitsukeru*

*osagashi*

Buying (2)

*omotome itadaku*

*koonyuu suru*

Utilization

*katsuyoo suru*
Location

haika suru

Borrowability

kashidashikanoo na jootai de aru

This range is not dissimilar to that of shoseki, but information occupies the major placings, and new semantic areas relate to utilization, location and availability for borrowing.

8 references of tosho, as an independent word or in composites, occurred in co-ordinate contexts involving sets of nouns. Partner terms were zasshi (3), shiryoo (2), bideo (2), shinbun. Again, zasshi is shared with shoseki but, as is clear in particular from the presence of shiryoo (and the absence of CD and DVD), tosho appears to involve greater emphasis on role as an information resource, irrespective of commercial factors.

Like shoseki, tosho appears in a large number of composite formations (including everyday items such as toshokan), and here we consider compounds, as for shoseki above. Tosh-o compounds in the corpus are predominantly of modifier-head structure: the commonest by far is toshomokuroku, followed by (complement-predicate) toshokensaku. -toso compounds with more than one occurrence in the corpus are shinkuntosho, shinchakutosho; suisentosho; sankootosho; nihongotosho; rokuuntosho.

2.3 Shomotsu

The corpus contained 42 references of shomotsu. 15 involved predicates taking shomotsu as subject or object argument, as listed below.

Influence (3)

eikyooryoku o motta
seiyyo o hizuuta
sehai o kaeta

Message (2)
Shomotsu occurs with a semantically relatively wide range of predicates: like shoseki and tosho they can be searched for and bought, but more often they are things which carry influence, contain a message, or are read. The collocational evidence suggests that the focus of shomotsu is commonly
on (significant) content.

In contrast to shoseki and tosho, shomotsu occurs in only one co-ordinate context in the corpus, paired with zasshi. It also appears to form far fewer composite expressions. Shomotsu- compounds were shomotsubunka (2 occurrences), shomotsukenkyuu, shomotsuseisan (as a historical process), and shomotsusoooshoku: -shomotsu compounds kanrenshomotsu, kakushushomotsu, ruijishomotsu, and amerikabijinesushomotsu. In addition to content area, these compounds focus on books as a cultural product.

2.4 Summary

While these three items all relate semantically to hon, textual tendencies indicate that they do so from different perspectives: shoseki commonly focuses on hon as a commercial product, tosho as an informational resource, shomotsu as a bearer of significant content and as a cultural product. Shoseki and tosho (but not shomotsu) are also shown to share collocations relating to information provision and search.

An additional semantic property of tosho, referred to in Kindaichi et al’s dictionary treatment above, is its status as a collective term (i.e. its English translation equivalent is books rather than book). The corpus contained one co-occurrence of shoseki and tosho, where the two terms clearly contrast in this respect. The discussion concerns a database consisting of a catalogue of books in Japanese (nihongotosho no mokuroku) on a particular topic, and includes the sequence Shikashi tosho ni wa kakushoseki no shozaiyoooh ga kakete ita: within collective tosho, individuated reference is made by means of shoseki.

3 Conclusion

Our aim here has been to demonstrate the value of collocational studies in the investigation of lexical meaning and use. To this end we have taken up two groups of related Japanese items, representative of ‘near-synonyms’ which pose problems for learners of the language, and examined collocational evidence of different kinds from a small corpus of attested examples.

Despite its limited scale — and we have in any case focused only on
selected collocational aspects of the corpus — we believe that our treatment has succeeded in revealing some significant collocational tendencies and in producing information on natural lexical use which goes beyond existing standard dictionary descriptions.

Native speakers of all languages control the meaning and use of a vast number of lexical items and, as noted earlier, they acquire this mastery on the basis of innumerable encounters with these items used in real contexts. By investigating linguistic aspects of these contexts, collocational studies can expand the range of information available to non-native learners in their own quest for lexical competence.

References
Cobuild 2001: Collins Cobuild English Dictionary for Advanced Learners
Miyajima T 1988 'Meishi, keiyooshi no buntai to imi.' Kokugogaku 154.

バックハウス、A. E. (留学生センター教授)