Homograph Disambiguation Using Formal Concept Analysis

L. John Old

School of Computing, Napier University, Edinburgh, UK
j.old@napier.ac.uk

Abstract. Homographs are words with identical spellings but different origins and meanings. Natural language processing must deal with the disambiguation of homographs and the attribution of senses to them. Advances have been made using context to discriminate homographs, but the problem is still open. Disambiguating homographs is possible using formal concept analysis. This paper discusses the issues, illustrated by examples, using data from Roget's Thesaurus.

Keywords: Type-10 chains, partitions, components, Roget's Thesaurus, plus operator, word fields, neighbourhood lattices.

1 Introduction

We use formal concept lattices [12] to extract and visualize ambiguous words (homographs) and senses from a lexicon, then use the results to identify whether the ambiguity of the words was resolved by partitions in the lattices. We then compare the results with previous attempts to disambiguate such words using Type-10 components [2], and analyse and discuss identified exceptions.

2 Homographs

Homographs[1] are words with identical spellings but different origins and meanings. These differences are made explicit in lexicons using headword numbers. The senses of a word are then identified under each headword.

1 bat
n. 1. bat [a club]; 2. bat [sports equipment]
vb. bat [to hit using a bat]
2 bat
n. bat [a flying mammal]
...

Also used are etymology (word history) and part-of-speech such as noun (n.) and verb (vb.), as in the example above. It is the lexicographer's judgment as to how many

[1] In linguistics the hypernym of homograph is homonym (meaning same name) and includes homophones, words of different meaning but which sound the same when spoken. An example is bore and boar.

B. Ganter and L. Kwuida (Eds.): ICFCA 2006, LNAI 3874, pp. 221–232, 2006.

divisions are made of a headword. The goal is to disambiguate words that look the same but are semantically distinct.

Headwords in lexicons may frequently be found differentiated numerically though they have a common etymological ancestry. In such systems *bat vb.* in the example above would be given a listing as a separate headword. This paper uses a strict definition of a homograph: that two or more words, though spelled the same, have different etymologies. For example, contract (with an emphasis on the first syllable, and meaning an agreement or to make an agreement) and contract (with an emphasis on the second syllable, and meaning to reduce in size), both derive from L. *contractus*, pp. *contrahere*, to draw together. Though these words have quite different meanings, they are not homographs under a strict definition, but cognates (related by descent from the same roots in some ancestral language). Bat1 (club) and bat2 (mammal) in the given example are indeed homographs under the strict definition. Bat1 is derived from L. *battuere* to beat (and, incidentally, cognate with battle and combat), while bat2 comes to English most likely from Old Swedish *nattbakka* (via early Viking settlers [13]).

All words used as examples in this paper were disambiguated etymologically by using their semantic roots. Apart from rare loan words borrowed from non-Indo-European language, the semantic roots used are all Indo-European (IE) roots. Note that all IE roots are hypothetical. They are sometimes referred to as proto-Indo-European roots and prefixed by a *, and are all derived by comparing words and their senses from Indo-European languages such as Latin, Greek and Sanskrit; and modern Icelandic, Hindi, Russian and Iranian (among others).

Note also that more than seventy homographs have more than two expressions. As an example, *pa* has three:

Homograph	**Meaning**
Pa 1	Protactinium
pa 2	father
pa 3	Maori stronghold

Pa 1 is an abbreviation of a chemical element; pa 2 is an English dialect word synonymous with father; and pa 3 is a foreign loan word. Though homographs with more expressions could be assumed to cause greater difficulties for disambiguation, usually they are rare words with low polysemy and generally easier to differentiate. It is the highly polysemous words (those with many sense variations) that provide the greatest difficulty for disambiguation.

3 Research

Natural language processing (NLP) must deal with the disambiguation of homographs and the attribution of senses to them. Advances have been made by employing context and systems such as N-gram taggers, Bayesian classifiers, vector space models with neural networks, and decision trees ([15],[16]) to discriminate homographs, but the problem is still open. Work has also been done using statistical models of Roget's Thesaurus categories to disambiguate word senses [14]. Treating words as objects and their senses as attributes, identifying words with disjoint sets of senses is possible using formal concept analysis (FCA) [17]. The assumption made and tested in this paper is that this is also applicable to differentiating homographs.

This research manually identified approximately 600 homographs using the strict definition. These were used as a benchmark to compare two partitioning methods, Type-10 chains ([2], [5], [9], [10]) and FCA lattices [12]. The test data was derived from Roget's Thesaurus [1], a semantic dictionary organized by concepts rather than words.

552 homographs were identified in RIT. Of those homographs, 179 either occur in senses by themselves or are the sole representatives of a homograph (the ambiguous partner does not occur in Roget's Thesaurus). For example Nice, a homograph of nice (likeable) appears only in an RIT list called Principle Cities of the World.[2] Not sharing any senses with other words, such homographs are already partitioned and cannot participate in Type-10 chains. Nice could also be easily differentiated using capitalization. The remaining 373 homographs are potentially ambiguous, and eligible for use as test words. These test words were used to compare the results of Type-10 components with neighbourhood lattices regarding effectiveness of discrimination of homographs.

Note that the goal here is to efficiently separate instances of homographs for disambiguation (lead1 from lead2), not to automatically identify and classify all instances of a particular homographic word together (all instances of lead1 together). The latter would indeed be desirable, but at this point entails further problems not yet solved. In other words, it is not yet possible, while partitioning homographs, to automatically group together all instances or senses of a particular word that is a homograph of another word, into a single partition.

4 Definitions

String a sequence of characters representing a word or homograph; an entry in RIT.

Word a disambiguated string, or entry, in RIT. Lead (lead1, guide, not follow) is one word; lead (lead2, the metal) is another, different, word.

Homograph a string for which there are two or more words with the same spelling, but with different etymological origins or roots. Lead1 derives from an Indo-European root, LEIT-2, meaning to guide; lead2 derives from an Indo-European root, EL-1, meaning red (from the colour of the oxide of lead, also known as red lead and sometimes used in primer paint).

Sense a set of words that share a particular meaning or concept; also known as a Synset; (the strings in this set are known as synonyms).

Entry a particular sense of a word; a particular word in a Synset.

Polysemy the number of senses a word has; number of entries representing a particular word in RIT. *Over*, for example, has a polysemy of 22; it can be found listed in 22 senses in RIT.

Paragraph a set of synsets of semantically-related concepts (and part-of-speech) , grouped together in RIT.

Category a set of paragraphs of semantically-related notions, grouped together in RIT.

[2] A list in Roget's Thesaurus is a special case where a set of senses each consist of a single word. Each entry in the list is viewed as a separate, but related, concept contributing to the list topic.

Figure 1 shows a sample of the structure of RIT using the word *over*. This represents one sense of over, and this instance of over is one entry in RIT. The entry occurs in Synset 40:10:1, read as RIT Category number 40; Category name, Addition; Paragraph 10; Synset 1. Each of the adjacent synonyms in Synset 40:10:1 is also an entry; a string representing one sense of that word.

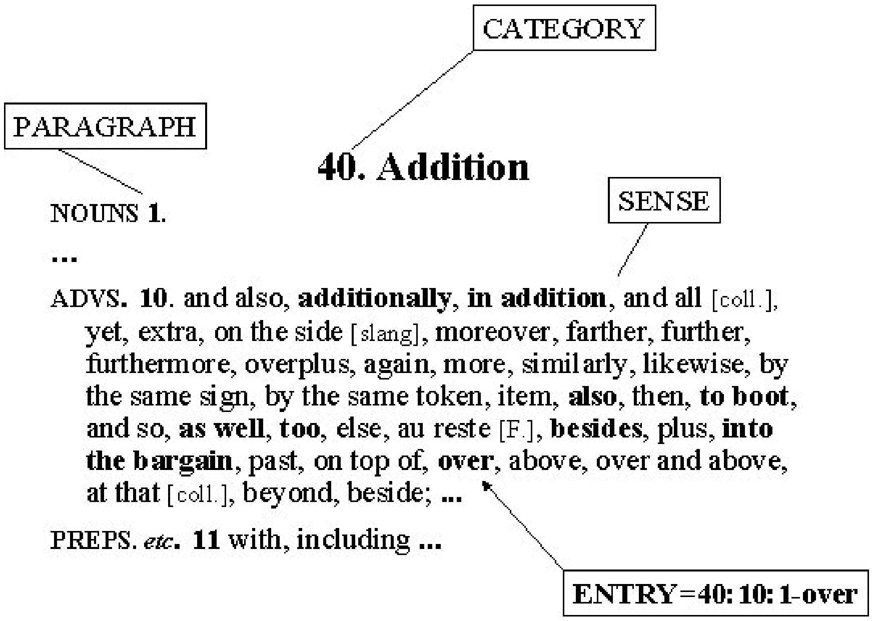

Fig. 1. An example of an entry in Roget's Thesaurus. The word *over* occurs as an entry in Category 40, Paragraph 10, Synset 1. (Bold-type synonyms are considered the most representative words of the sense).

5 Formal Concept Analysis

Several researchers have used so-called neighbourhood lattices to visualize parts of Roget's thesaurus. A semantic neighbourhood is similar to a word field (a set of semantically related words), but also includes the set of the shared senses of the words. A formal context built from a semantic neighbourhood takes the words of a neighbourhood as formal objects and their corresponding senses as formal attributes. The original formalization of neighbourhood lattices was suggested by Wille in an unpublished manuscript. Priss [7] defines neighbourhood lattices as follows:

Instead of using the prime operator ($'$), the plus operator ($+$) retrieves for a set of objects all attributes that belong to at least one of the objects. Formally, for a set G_1 of objects in a context (G, M, I), $\iota^+(G_1) := \{m \in M \mid \exists_{g \in G_1} : gIm\}$. Similarly $\varepsilon^+(M_1) := \{g \in G \mid \exists_{m \in M_1} : gIm\}$ for a set M_1 of attributes. If two plus mappings are applied to a set G_1 it results in a set $\varepsilon^+\iota^+(G_1)$ (with $\varepsilon^+\iota^+(G_1) \supseteq G_1$) which is called the *neighbourhood* of G_1 under I. A neighbourhood of attributes is defined analogously. A neighbourhood context is a context whose sets of objects and attributes are neighbourhoods, such as $(\varepsilon^+\iota^+(G_1), \iota^+\varepsilon^+\iota^+(G_1), I)$. The resulting lattice is called a neighbourhood lattice.

Such lattices are used here to collect and display the senses and synonyms of a topic word. Though there is no limit to the number of times the plus operator can be applied, three times is sufficient to create the neighbourhood of senses with synonyms. The first iteration is to collect the senses of the topic word; the second is to collect the synonyms shared within those senses; and the third to collect the special senses of the synonyms (not shared by the topic word).

6 Type-10 Components

Type-10 chains and Type-10 components derive from the mathematical model of abstract thesauri (of which Roget's Thesaurus is one instantiation), developed by Bryan [2]. The elements in this model are strings and senses (sense definitions, or Synsets), and a relation between them. Bryan defined a series of chains linking entries by word associations, sense associations, or both. If a word appears in two different senses, an association between the senses is implied. If two different words share a sense, an association between the words is implied. The most restrictive, the Type-10 chain, is a double chain. This requires that for any two words sharing a sense, there exists a second sense that that also shares those two words, in order to participate in the chain. The two words plus two senses has been dubbed a *quartet*.

The Type-10 chain restriction is intended to ensure that links are not arbitrary, as happens when two senses are linked by homographs. The assumption is, for example, that there is no second word that accompanies both *lead* (the metal) and *lead* (not follow) for any pair of their senses. Figure 2 illustrates a simple example of RIT entries (the X's) forming quartets.

Type-10 chains are used to form partitions (equivalence relations), called here *components*, on sets of entries. Talburt and Mooney [10] derived all possible Type-10 chain components from the 200,000 entries in RIT in an attempt to automatically separate

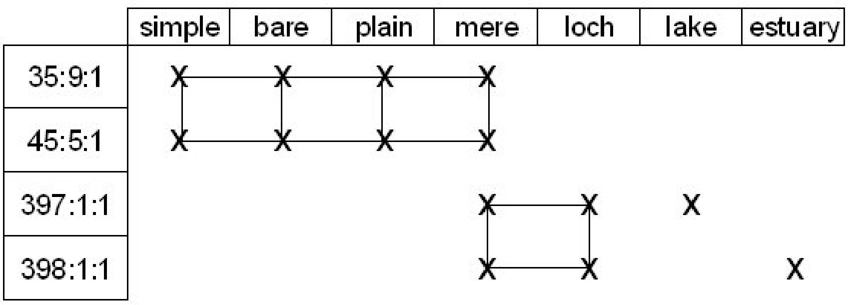

Fig. 2. Quartets formed by some of the synonyms and senses of *mere*. *397:1:1 / 398:1:1 / loch / mere* forms one quartet.

all homographs. The largest Component contained more than 22,000 entries, and more than 10,000 components were derived. The effectiveness of this method has never been tested against a complete set of homographs.

Jacuzzi [5] reproduced Talburt and Mooney's work but applied a further constraint: that a quartet can not participate in a component if it shares only one RIT entry with that component. This was because he observed that it was possible for a quartet to satisfy the Type-10 constraint, yet connect to other quartets by only one entry (a string identical to one or more other strings in the component, plus a sense shared by one or more other strings in the component).

Strictly speaking, Jacuzzi's derived components are not partitions because in splitting quartets the offending entry must be included in all derived child components. The components are no longer equivalence relations on the set of all entries in RIT. None-the-less Jaccuzzi's results were chosen for this study as they are more restrictive and the components are smaller and therefore less likely to combine homographs. The maximum sized Jacuzzi component is 1,490 RIT entries.

7 Neighbourhood Lattices

Mere is a homograph. Its main meanings are simple, pure (mere1), and sea, ocean, loch (mere2). Webster's New Collegiate Dictionary [13] defines mere2 as a sheet of standing water: POOL. Figure 3 shows the neighbourhood lattice of the undifferentiated word. The left hand side (the group to the left of the object *mere*) represents mere1. Apart from the concept labelled with *mere*, this forms a partition separating mere1 from senses of mere2. For simplicity, the attribute labels are not shown in this diagram. Two of mere1's attributes, for example, are 35:9:1 Smallness and 45:5:1 Simplicity, Noncomplexity.

Most of the synonyms of mere in Figure 3 are differentiated by senses that are peculiar to them. This causes a more complex structure than is needed to observe partitions between instances of the topic word (the homograph we are interested in).

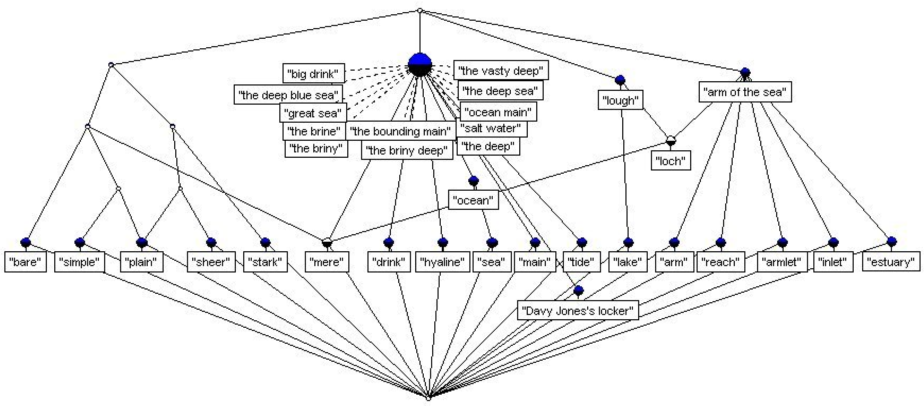

Fig. 3. Formal concept lattice of the semantic neighbourhood of mere (all synonyms and all senses (unlabelled) of the synonyms of mere)

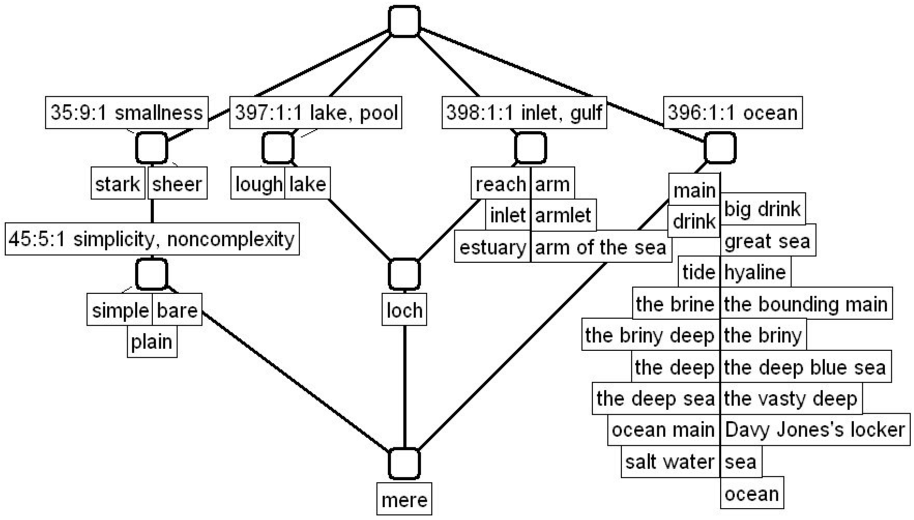

Fig. 4. Formal concept lattice of the semantic neighbourhood of mere (synonyms and senses of mere only)

Figure 4 shows a neighbourhood lattice of mere where the plus operator has been applied only twice. Here, the senses of *mere* alone are used as attributes. Mere1 is identifiable by the two senses 35:9:1 and 45:5:1.

If the top- and bottom-concepts of the lattice are removed from the lattice in Figure 4 we obtain three disjoint graph components; one dealing with mere1 and two dealing with mere2. This shows that there is no overlap of senses or synonyms of mere1 and mere2. The result is called the horizontal decomposition of the lattice and has been used amongst others, by Dekel and Gil [3] to identify component classes in the structure of legacy computer software. The collective lattice of partitions is called the horizontal sum [4]. Thus, Figure 4 is the horizontal sum of the meaningful components.

Figures 3 and 4 illustrate the separation of homographs of a word using the senses as formal attributes (and as differentiae). Of the approximately two-and-a-half-thousand entries in Roget's Thesaurus that represent homographs, all but 22 were found to be differentiable by this method. Those 22 ambiguous entries consisted of 10 homographs. The same entries were differentiable by the Type-10 partitioning (i.e. the 22 were also not differentiable by Type-10 components). The 22 undifferentiated entries (10 homographs) were found among just eight of the Jacuzzi components.

8 The Exceptions

The 10 problem homographs fell into three cases. Each case involved a word shared by each of the homographs (a synonym in common). This shared word provided a bridge between the homographs and prevented the formation of a partition. The 10 homographs

Table 1. Ambiguous homographs, their Indo-European roots and root meanings

Entry	IE Root	Root Meaning
brash 1	BHEL-2	swell, blow
brash 1	KAU-2	strike, hew
brash 1	RE-1	Backward
brash 2	BHREG-	break, breach
fell 1	P(H)OL-	Fall
fell 2	GHEL-2	shine, bright
light 1	LEGWH-	light, not heavy
lightsome 1	LEGWH-	light, not heavy
light 2	LEUK-	light, brightness
lightsome 2	LEUK-	light, brightness
post 1	STA-	stand
poster 1	STA-	stand
post 2	(A)PO-	away, off
poster 2	(A)PO-	away, off
press 1	PER-5	strike
press 2	GHESOR-	hand
rash 1	KAU-2	strike, hew
rash 1	RE-1	backward
rash 2	RED-	scrape, scratch, gnaw
set 1	SED-1	sit
set 2	SEKW-1	follow
set 3	N/A	Egyptian god
slug 1	SLAK-	strike
slug 2	SLEU-	sluggish, slow

were *brash, fell, light, lightsome, post, poster, press, rash, set* and *slug*. These are listed in Table 1, along with their Indo-European Roots and each roots' meanings[3].

The first case involved homographs that each shared a supposedly unambiguous (non-homographic) word as a synonym. As it happens these shared words had many, very diverse senses (25 in the first instance). One of those senses of the shared word overlapped with the meaning of one homograph, while a second sense of the word overlapped with the meaning of the second homograph. Such cases will be referred to as the ambiguous synonyms category.

As an example, press1 generally relates to printing or pressure, and is derived from an IE root, PER-5 (meaning, to strike); while press2 relates to drafting into military service or being at hand, and derives from GHESOR- (meaning, hand). Both press1 and press2 share *call* as a synonym. The sense of the word call shared with press1 has to do with calling *on* someone, as in: to pressure someone for money or a sale. The second sense of call, shared with press2, has to do with calling *up* someone (call up has one meaning of ordering someone to report for military duty). It is clear that these

[3] Brash1 comes from bold + rash1; rash1 is a variant of rush1 (hurry); rush1 and rash1 come from RE-1 (backward) + KAU-2 (strike, hew); and bold comes from BHEL-2 (swell, blow). Brash1, therefore, derives from the three IE roots BHEL-2 + RE-1 + KAU-2. Hence the three entries in the table for brash1, and the two entries for rash1.

Table 2. Category one of undifferentiated homographs involves ambiguous synonyms

Hom1	Meaning1	Hom2	Meaning2	Comp#
Press1	**call** on	Press2	**call** up	[VJ2 2478]
Set1	**head** for	Set2	**head** of (collection)	[VJ2 184]
Set1	**suit**, attune (to match)	Set2	**suit** (matching attr.s)	[VJ2 9323]

senses of two, otherwise distinct and unambiguous, homographs live close to each other in the semantic universe.

The other homograph in this category, *set*, has two instances of this problem. Considering that the word *set* has 53 senses, 51 of which were disambiguated, this is not a dismal result. The ambiguous entries in Roget's Thesaurus for the first category are listed in Table 2. The last column shows the component numbers from Jaccuzzi's Type-10 chain components.

Figure 5 illustrates the restricted neighbourhood lattice of the homographs of *press*. To reduce complexity the formal context was restricted to those synonyms that occur in more than one sense of press. The majority of senses belong to press1. Press2 has the three senses located to the extreme right. *Call* is a shared synonym of both press1 and press2, and prevents the formation of a partition between the two homographs.

The second category of undifferentiated homographs is identifiable by shared synonyms that are themselves ambiguous homographs. These are, again, highly polysemous words. They have a wide range of senses that allow them to overlap semantically, as synonyms, with many other strings-including the strings that constitute the instances of these topic homographs. Table 3 shows the second category.

Down2 is an expanse of rolling, grassy, treeless upland used for grazing (a moor) in the context of the undifferentiated RIT fell entry. It is often used in the plural, in this sense, as part of a place-name, as in: *Watership Downs*. A fell (fell2) is a type of flat

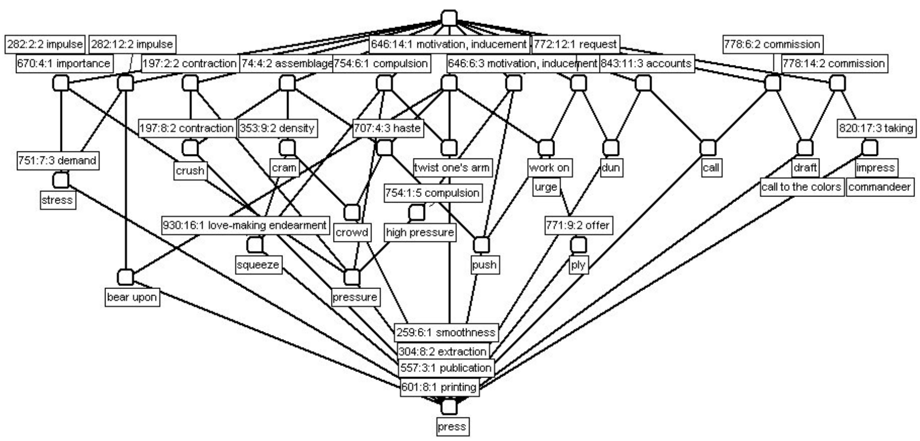

Fig. 5. Formal concept lattice of the restricted semantic neighbourhood of press (all synonyms sharing more than one sense, plus senses of press only). Press2 has the three senses located to the extreme right. Call is a shared synonym of both press1 and press2.

Table 3. Category two of undifferentiated homographs involves ambiguous homographs

Hom1	Meaning1	Hom2	Meaning2	Comp#
Fell1	fall, Down2	Fell2	moor, a Down2	[VJ2 4573]
Slug1	slowpoke, Poke1	Slug2	slog, Poke1	[VJ2 1501]

land, synonymous with moor. Down2 is also the adverb, *down* (as in: up, down, forward and backwards). It is an extension of this second persona of down2 which is found as a synonym of fell1 (to down something; to chop down or drop, as in: to fell a tree). Down1, incidentally, is the down derived from goose feathers.

Poke1 is an abbreviation of an Americanism, slowpoke [US], from cowpoke [US], a cowpuncher [US] or cowboy. Punching cows involved prodding (or poking) them with poles to make them enter railroad cars. Slug1 is the slimy invertebrate and slowpoke means one who moves slowly. Slug2 means hit; strike. (Poke2, not listed, is a pocket, as in: a pig in a poke).

Category three involves three cases where synonyms that are differentiable homographs co-occur with related homographs that are themselves differentiable homographs, but that match with homographs accompanying their homographic nemeses. The consequence is that there is a catch-22; each homograph relies on the other to differentiate it. Because they have identical spellings, and their cohort homographs have identical spellings, they match in a rare undifferentiated quartet. Referring to Table 4, light1 (not heavy) and light2 (bright), both have *lightsome* as a synonym. Lightsome1 means weightless (also cheerfulness and caprice) while lightsome2 means full of light. Lightsome (1 & 2) is an infrequent and archaic word in English.

Likewise, *brash* and *rash* can both mean impulsive or indicate a collection of red spots that are symptoms of a disease. And *post* or *poster* can both mean put up an announcement and also be the names for the role of a mailman. Brash2, poster1, and poster2 are all rare words in English.

Note that in category three, lightsome, poster and rash are equally ambiguous to light, brash, and post. So this represents six, not just three, instances of undifferentiated homographs. Table 4 was restricted to emphasizing light, brash, and post, for brevity.

Table 4. Category three of undifferentiated homographs involves other undifferentiated homographs

Hom1	Meaning1	Hom2	Meaning2	Comp#
Light1	weightless, Lightsome1	Light2	bright, Lightsome2	[VJ2 7393]
Brash1	impulsive, Rash1	Brash2	disease, Rash2	[VJ2 2183]
Post1	(put up a) Poster1	Post2	mailer, (a) Poster2	[VJ2 8564]

9 Discussion

It is well known by linguists that words with otherwise distinct etymologies can influence each other to trade meaning and blend, and eventually come to have the same connotations. This may well have happened with press1, press2, light1 and light2, and possibly some of the other homographs unsuccessfully disambiguated here by either the FCA lattices or the Type-10 components methods. Also, many of these ambiguous synonyms and homographs are highly polysemous, giving ample opportunity for instances of semantic overlap.

Finally, in some of these cases one of the senses or one of the words is rare. It should not be surprising that these are difficult to disambiguate. Languages tend to discard or modify ambiguous words, but rare instances have less opportunity for scrutiny. In fact many of the successfully disambiguated homographs suggested a pattern. That if a homograph came from a specialty area, such as a branch of science, it was more likely to have a matching common word as a homograph. An example is abbreviations for chemical elements. Be: Beryllium; He: Helium; As: Arsenic; In: Indium; At: Astatine.

It may appear to be a contradiction that both highly polysemous and rare-sense words should both contribute to ambiguity, but there appears to be a balance required for disambiguation in language-sufficient context to differentiate but not so much as to cause confusion.

10 Conclusion

We have compared more than two-and-a-half-thousand semantically ambiguous entries in Roget's Thesaurus using two methods, FCA neighbourhood lattices and Type-10 chain components. The ambiguity amongst entries was caused by homographs-words of identical spelling but with different origins and meaning. We conclude that, given a lexicon and set of homographs in common, FCA neighbourhood lattices can discriminate homographs as well as Type-10 chain components. Furthermore, while Type-10 components may contain up to 1,500 thesaurus entries, semantic neighbourhoods are constrained to the senses and synonyms of the topic word. Consequently, the partitions formed around homographs using FCA make the data more tractable and human-accessible.

Ten of the 373 homographs used in this study had senses that were undifferentiable by either of the two methods. These cases involved senses where homographs of completely different origins (by definition) overlapped semantically via words in common. These words-in-common were other homographs in all but three instances. The failed instances involve very rare or very common (highly polysemous) words, and may represent the boundaries for discrimination of homographs. They may also indicate the range and combination of frequency and rarity necessary to disambiguate polysemous words in human conceptual processing.

Future work should examine the effectiveness of combining previously documented methods of homograph disambiguation with FCA neighbourhood lattices to disambiguate homographs with a view to improving effectiveness. Furthermore, while neighbourhood lattices are effective at partitioning senses of homographs, at the same time they may partition senses within the set of senses. For completeness, a method should be developed which classifies together all of the senses of any homograph to which the partitioning method is applied.

Acknowledgements

The lattices used in this paper were developed using Conexp [17] and Anaconda [11]. The electronic version of Roget's International Thesaurus was used with kind permission from Dr W. A. Sedelow Jr. and Dr S. Yeates Sedelow.

References

1. Berrey, L. (Ed.). (1962). *Roget's international thesaurus (3rd ed.).* New York: Crowell.
2. Bryan, R. M. (1973). Abstract thesauri and graph theory applications to thesaurus research. In S. Y. Sedelow (Ed.), *Automated language analysis, report on research 1972-73* (pp. 45-89). Lawrence, KS: University of Kansas.
3. Dekel, U., and Yossi, G. (2003). Revealing class structure with concept lattices. *The Tenth Working Conference on Reverse Engineering,* IEEE Computer Society Press, pp. 353-365.
4. Ganter, B., and Wille, R. (1999). *Formal concept analysis: Mathematical foundations.* Berlin-Heidelberg-New York: Springer. ISBN 3-3540-62771-5.
5. Jacuzzi, V. (1991, May). *Modeling semantic association using the hierarchical structure of Roget's international thesaurus.* Paper presented at the Dictionary Society of North America Conference, Columbus, Missouri.
6. Miller, G., Beckwith, R., Fellbaum, C., Gross, D., Miller, K., and Tengi, R. (1993). *Five papers on WordNet.* Technical Report. Princeton, N.J: Princeton University.
7. Priss, U. (1996). *Relational concept analysis: Semantic structures in dictionaries and lexical databases.* (Doctoral Dissertation, Technical University of Darmstadt, 1998). Aachen, Germany: Shaker Verlag.
8. Priss, U., and Old, L. J. (2004). Modelling lexical databases with formal concept analysis. *Journal of Universal Computer Science*, *10*(8), 967-984.
9. Sedelow, S.Y. (1991). Exploring the terra incognita of whole-language thesauri. In R. Gamble and W. Ball (Eds.), *Proceedings of the Third Midwest AI and Cognitive Science Conference* (pp. 108-111). Carbondale, IL: Southern Illinois University.
10. Talburt, J. R., and Mooney, D. M. (1990). An evaluation of Type-10 homograph discrimination at the semi-colon level in Roget's international thesaurus. *Proceedings of the 1990 ACM SIGSMALL/PC Symposium*, 156-159.
11. Vogt, F. (1996). *Formale Begriffsanalyse mit C++: Datenstrukturen und Algorithmen.* Springer-Verlag, Berlin-Heidelberg, 1996, ISBN: 3-540-61071-5.
12. Wille, R. (1982). Restructuring lattice theory: an approach based on hierarchies of concepts. In I. Rival, (Ed.), *Ordered sets* (pp. 445-470). Dordrecht: Reidel.
13. Woolf, B. H. (Ed.). (1976). *Webster's new collegiate dictionary*, G. & C. Merriam Company, Springfield, Massachusetts.
14. Yarowsky, D. (1992). Word-sense disambiguation using statistical models of Roget's categories trained on large corpora. *Proceedings of 14th International Conference on Computational Linguistics, COLING-92.* Nantes, pp. 454-460.
15. Yarowsky, D. (1996). Homograph disambiguation in text-to-speech synthesis. In J. van Santen, R. Sproat, J. Olive, and J. Hirschberg, (Eds.), *Progress in Speech Synthesis*: Springer, New York, 1996.
16. Yarowsky, D. (2000). Word sense disambiguation. In R. Dale, H. Moisl and H. Somers (Eds.) *The Handbook of Natural Language Processing.* New York: Marcel Dekker, pp. 629-654.
17. Yevtushenko, S. (2005). Conexp (Concept Explorer) 4-Beta. *SourceForge.net.* Available via http://sourceforge.net/projects/conexp