Kobe University Repository : Kernel

PDF issue: 2026-11-03

Lexical Priming and Asian Learners of English

Ooi, Vincent

(Citation)
Learner Corpus Studies in Asia and the World,1:31-41

(Issue Date)
2013-03-23

(Resource Type)
departmental bulletin paper

(Version)
Version of Record

(JaLCDOI)
https://doi.org/10.24546/81006673

(URL)
https://hdl.handle.net/20.500.14094/81006673

KOBE UNIVERSITY

Lexical Priming and Asian Learners of English

Vincent OOI
National University of Singapore

Abstract

In this paper, the notion of lexical priming (Hoey 2005) for achieving 'nativeness' and 'naturalness' will be mainly investigated using a Corpus Query System known as the Sketch Engine which automatically produces 'Word Sketches' (Kilgarriff 2009). The application of Word Sketches will be demonstrated by investigating the top two words (i.e. *part-time job, smoking*) that make up the topics written by college/university students in the International Corpus of Asian Learner English (ICNALE). Such Word Sketches can be used to examine the statistically 'salient' verbs and other grammatical/collocational information associated with these two top nouns in a few of the learner corpora from countries such as Japan, Singapore, and Hong Kong. A further comparison can be made with the native-learner (ENS1) and native-speaker (ENS2) corpora gathered for the ICNALE project.

Keywords

Lexical priming, ICNALE project, Learner corpus, Varieties of English, Wordsketch, Naturalness, Nativeness

I Introduction

This paper discusses the central notion of 'lexical priming' (Hoey 2005) and examines aspects of its acquisition among Asian learners of English, The term 'Asian learners' needs clarification in terms of its reference to geography, nationality, or ethnicity (leaving aside the philosophical notion of what constitutes 'Asian'). In this paper, we utilize the ICNALE corpus (Ishikawa 2011) as the main source of linguistic evidence. Its associated website, http://icnale.por.jp/login.php, offers standard features of a concordancer, i.e. Keyword-in-Context view, Collocation, Wordlist, and Keywords. Therefore, 'Asian learners' is operationally used to refer more to the nationality of learners in those countries that have participated in the project.

Besides the highly motivating online website with its corpus tools (just referred to), another Corpus Query System known as the Sketch Engine (Kilgarriff 2009, Kilgarriff and Tugwell 2002) will be used to produce 'Word Sketches'. A wordsketch is 'a one-page,

automatic, corpus-derived summary of a word's grammatical and collocational behaviour' that saves both the researcher and the learner time and effort in calculating the necessary frequencies of usage.

In this paper (and oral presentation), the analysis will therefore be largely lexical in nature. Owing to constraints of space, this paper version will be limited to a few examples which will nevertheless make clear the direction of the presentation. Also, while we use corpus evidence as the main source of linguistic evidence, let us not forget that we should also emphasize the central social sciences notion of 'triangulation' in which findings are corroborated based on two or more types of linguistic evidence: introspection, citations, surveys, dictionaries, grammar books etc.

II Lexical priming

Why should we focus on 'lexical priming'? The reason for this is that lexical priming is centrally related to 'naturalness' and 'nativeness' – the twin notions that underscore standard English as the goal of language learning and teaching. Achieving 'naturalness' (while eschewing 'clumsiness') is equally as – if not more -- important, in terms of how 'native' the sentence or linguistic expression might be. In turn, 'nativeness' has to be contextualized in terms of the target native variety, of which the two main benchmarks are standard British English and standard U.S. English.

In recent years, there has been a growing realization that the key to achieving these objectives of naturalness and nativeness is a good grounding in linguistic primings. Let us introduce a few basic assumptions underscoring the notion of lexical priming (Hoey 2005):

(p. 2): "[Lexical priming] is concerned, in part, with how naturalness is achieved and how an explanation of what is natural might impinge on explanations of what is possible."

(p. 11) "Primings may crack, and one of the causes of cracking is education. If, for example, a word is primed for someone as collocating with a particular other word and a teacher tells that person that it is incorrectly primed (e.g. *you* and *was*) the result is a potential crack in the priming. Cracks can be mended either by rejecting the original priming or by rejecting the attack on the priming."

(p. 14) "A corpus, whether general – like the British National Corpus or the Bank of English – or specialised...represents no one's experience of the language....the computer corpus cannot tell us what primings are present for any language user, but it can indicate the kinds of data a language user might encounter in the course of being primed. It can suggest the ways in which priming might occur and the kinds of feature for which words or word sequences might be primed. In other words, it can serve as a kind of laboratory in which we can test for the validity of

claims made about priming."

(p. 183) "..the learning of a second language (L2) is necessarily a very different experience from learning a first one (L1)...In the first place, when the vocabulary of the first language is primed, it is being primed for the first time. When the second language is learnt, however, the primings are necessarily superimposed on the primings of the first language."

(pp. 185-186) "...priming is the result of a speaker encountering evidence and generalising from it....Language teaching materials and language teachers can provide essential shortcuts to primings...[this range includes] usage notes, drilling exercises, texts or tapes with repeated instances of a word sequence, collocational observations and illustrations or just drawing a class's attention to a feature all may speed up the process, as may judicious use of corpus-based monolingual dictionaries and corpus-based descriptive grammars."

(p. 186) "[However], unhelpful primings may result from a textbook's overemphasis on certain features of the language..or on its fabricated illustrations of grammatical points. At best, unhelpful primings will result in cracks in the priming when the learner encounters authentic instances of the language away from the teaching context. This may lead to insecurity or distrust of the value of what has been learnt in the classroom. At worst, [unhelpful primings] may inhibit the development of helpful primings and stunt language growth."

As a concrete illustration of what lexical priming entails, let us consider the following sentences in Hoey (2005: 5):

In winter Hammerfest is a thirty-hour ride by bus from Oslo though why anyone would want to go there in winter is a question worth considering.

Through winter, rides between Oslo and Hammerfest use thirty hours up in a bus, though why travellers would select to ride there then might be pondered.

As you may guess, the first sentence (from Bill Bryson's travel book *Neither Here Nor There*) is natural while the second sentence (deliberately made up by Hoey) is clumsy. But why is so? A close corpus inspection will show that the chances of co-occurrence for words such as *in* and *winter* occur more frequently than *through* and *winter*; similarly *by bus* occurs much more frequently than *in a bus*. One can continue with this reasoning for the other sequences in the sentence.

Simply put, we have our respective 'mental frequencies' for various types of phonetic, grammatical and discoursal knowledge that prime us to use language in particular ways.

In turn, our collective 'frequencies' are aligned to the corpus that is said to be

maximally representative of the linguistic question that we seek to answer. In order to discover these frequencies more precisely, we need access to appropriate corpora and software tools that help us in calculating textual patterns that would take too long for us to analyze using a 'pen-and-paper' approach. Such textual patterns afforded by corpus tools can include frequency lists, concordances, semantic annotations, and 'Word Sketches.'

III The learner corpus and the ICNALE project

At first sight, a learner corpus seems to be an unsavoury object, given that it may be linguistically error-ridden and contains unnatural expressions. However, Flowerdew (2009: 343) reminds us that 'frequency data drawn from learner corpora can show how learners may over- or under-use certain patterns of the target language' and that such data allows us to identify 'discrepancies between different stages of interlanguage and between L2 and L1 users on the basis of "errors".'

While there are proponents who emphasize the spoken component of learner language first (e.g. Tarone and Swierzbin 2009), Ishikawa (2011: 3) also appropriately focuses on the written component for the ICNALE project in order to achieve similar central objectives: i) studies of "interlanguage" and ii) identifying learner errors, overused and underused patterns. 'Interlanguage' is a concept embraced perhaps especially by EFL countries which construe their own production of this second/foreign language as neither their first language nor 'perfectly native'. The comparison of a learner corpus with an equivalent native one (and, for that matter, between various learner corpora collected in a similar manner) is also a worthwhile enterprise. Learner errors can be identified so that language teachers and students can focus on reducing their production and reception of these features more commonly occurring in their respective linguistic environments.

The International Corpus Network of Asian Learners of English (Ishikawa 2011) is a one-million-word learner corpus open to the public. According to the updated website (http://language.sakura.ne.jp/s/icnale.html), it collects essays written by 'college students in 10 Asian countries and areas as well as by English native speakers.' The topics are controlled for "(non-)smoking" and "part-time job".

IV The Sketch Engine/WordSketch

We now turn to the Sketch Engine (Kilgarriff 2009, Kilgarriff & Tugwell 2002), 'a corpus tool which takes as input a corpus of any language and a corresponding grammar patterns and which generates Word Sketches for the words of that language. A Word Sketch is a 'one-page, automatic, corpus-derived summary of a word's grammatical and collocational behavior.' The idea of the Word Sketch is motivated in

part by those learners who find it difficult and confusing to inspect concordances manually. As Kilgarriff (2009) puts it, many language learners are 'scared off' by concordances too difficult to analyse.

Support for the automatic analysis of concordances also comes from Cheng (2012) who cites the following responses from corpora-list subscribers in April 2008:

i) "Concordances are too difficult for learners";

ii) "Data-driven learning is inaccessible to many teachers for lack of know-how or resources";

iii) "Data-driven learning is inaccessible to many learners, as it presupposes an introspective, reflective approach to study, using new and relatively difficult technology, by learners and teachers who may not be highly motivated.

In this regard, the Sketch Engine allows the entire input text to be first annotated and relevant frequencies automatically derived from the corpus. For instance, in the case of *smoking* (which constitute one of the two topics that students write about in the ICNALE corpus), a partial Word Sketch for the construal of the term for native speakers in the Collins Cobuild Database ('WordBanks Online') of approximately 550 million words (see Figure 1) is as follows:

smoking **WordBanks Online: English freq = 15916**

object_of	3911	2.7	subject_of	2165	2.3	adj_subject_of	301	2.6	a_modifier	922	0.6	n_modifier	1296	1.7
quit	320	9.27	ruin	18	5.83	harmful	22	6.97	passive	396	11.22	cigarette	278	8.8
ban	476	9.22	cause	153	5.43	addictive	9	6.8	maternal	17	7.3	dope	13	7.12
stop	660	7.68	harm	10	5.24	prevalent	5	5.5	indoor	14	6.24	cannabis	24	6.86
outlaw	39	7.63	shorten	5	5.13	dangerous	15	3.47	second-hand	6	6.14	tobacco	41	6.81
prohibit	55	7.3	ban	12	4.01	bad	51	3.32	teen	9	5.99	opium	9	6.46
restrict	38	6.34	kill	61	3.99	responsible	9	2.74	careless	6	5.87	stop	54	6.17
permit	41	6.17	double	7	3.71	easy	12	1.98	parental	5	5.1	cigar	10	6.04
start	361	6.14	lower	6	3.7	hot	6	1.78	teenage	11	4.97	marijuana	15	5.96
overtake	16	6.12	damage	8	3.69	legal	5	1.36	heavy	38	4.63	pot	33	5.8
curb	17	6.07	affect	24	3.68	important	7	0.21	mum	5	4.62	workplace	10	5.42

Fig. 1 Word Sketch for *smoking* (n) (from WordBanks Online)

The top 5 salient verbs that precede *smoke* tend to be the following: *quit* (320 instances), *ban* (476 instances), *stop* (660 instances), *outlaw* (39 instances) and *prohibit* (55 instances). These results are not dissimilar to the native-speaker corpus in the ICNALE project, of which the top verbs include *ban, allow* and *prohibit* (Fig. 2)

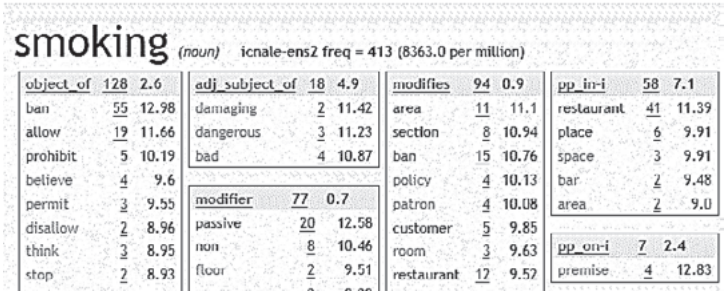


Fig 2. Word Sketch for *smoking* (n), from the ENS2-Native speaker ICNALE corpus)

V Results and Discussion (for verbs that precede *smoking* and *part-time job* in non-native Asian countries)

Extending the discussion on *smoking* in the preceding section, the following (Figures 3,4 and 5) show the respective frequencies for the salient verbs preceding *smoking* in the Singapore, Hong Kong and Japanese components of the ICNALE project respectively.

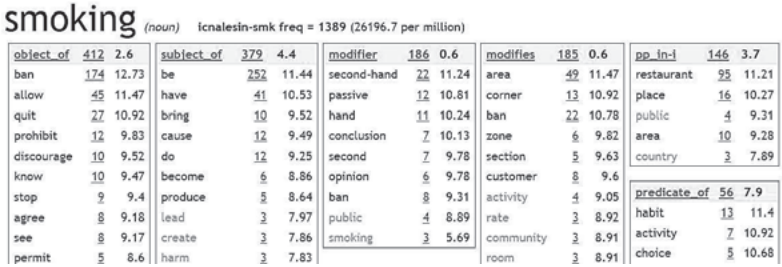


Fig 3. Word Sketch for *smoking* (n) (from the Singapore ICNALE corpus)

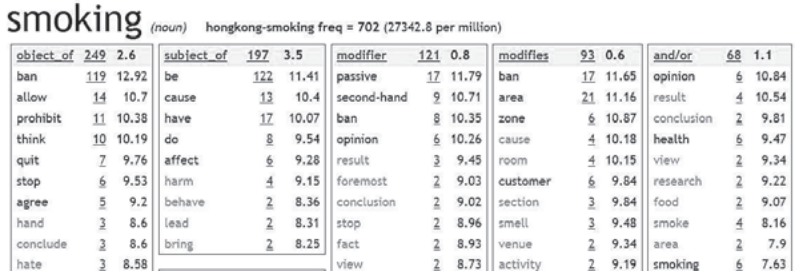


Fig. 4 Word Sketch for *smoking* (n) from the Hong Kong ICNALE corpus)

smoking *(noun)* **icnalejapan-smk freq = 3015 (30541.5 per million)**

object_of	937	2.7
ban	302	12.4
think	122	11.55
stop	85	11.31
prohibit	42	10.38
like	36	10.02
hate	30	9.84
forbid	18	9.22
know	18	9.16
say	16	8.91
allow	14	8.85

subject_of	819	3.3
be	509	11.47
have	65	10.01
make	33	9.62
do	50	9.62
cause	18	9.28
give	15	9.03
affect	14	9.01
harm	10	8.47
annoy	5	7.56
lead	4	7.29

modifier	402	0.7
passive	116	12.73
ban	19	10.43
active	12	9.87
secondhand	6	8.87
second	6	8.75
example	6	8.66
third	5	8.59
opinion	7	8.57
first	5	8.54
day	5	8.48

modifies	715	1.2
area	132	11.78
seat	137	11.61
people	156	10.57
person	39	10.31
section	27	10.09
room	27	9.95
place	27	9.25
space	16	9.15
man	11	8.79
smell	10	8.22

and/or	185	1.0
example	5	9.26
addition	4	9.21
instance	4	9.15
opinion	6	8.99
other	4	8.91
reason	8	8.87
nonsmoker	4	8.4
health	5	7.79
place	5	7.47
smoker	9	7.37

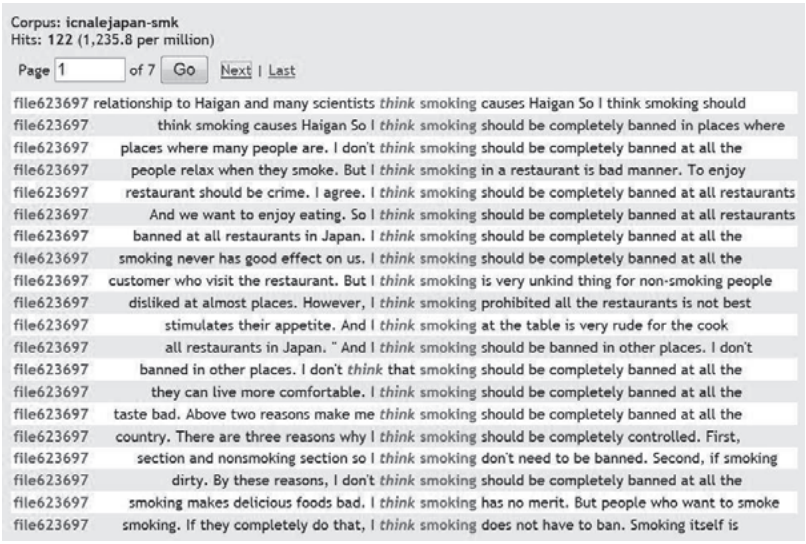
Fig 5 Word Sketch for *smoking* (n) (from the Japan ICNALE corpus)

Arguably, greater 'nativeness' can be said to include the use of verbs such as *quit*, *ban* and *prohibit* before the term *smoking*. The use of the verb *think* that occurs in the top few positions for some of the corpora is intriguing, though, and a closer inspection of the term is as follows (Figures 6 and 7):

Corpus: hongkong-smoking
Hits: 10 (389.5 per million)

file623705	and city's image can be improved. I *think* smoking shouldn't be banned at all restaurants
file623705	these places. On the other hand, I *think* smoking should be banned in all restaurants, as
file623705	right if they abide by the law. I *think* smoking should completely ban at all restaurant
file623705	air-circulating system. Therefore, I *think* smoking should completely ban at all restaurant
file623705	and that is awful. All in all, I *think* smoking should totally be banned in public places
file623705	restaurants is having many advantages. I *think* smoking should be prohibited at all area in the
file623705	been in the hot seat recently. I *think* smoking should be completely banned at all restaurants
file623705	these places. On the other hand, I *think* smoking should be banned in all restaurants, as
file623705	the measure imposed. To conclude, I *think* smoking in restaurants completely banned is a good
file623705	meaningless. In conclusion, I do *think* that smoking should be banned at all restaurants in

Fig 6. Edited concordance of *smoking* (n) from the Hong Kong ICNALE corpus)



Corpus: icnalejapan-smk
Hits: 122 (1,235.8 per million)

Page 1 of 7 Go Next | Last

file623697	relationship to Haigan and many scientists *think* smoking causes Haigan So I think smoking should
file623697	think smoking causes Haigan So I *think* smoking should be completely banned in places where
file623697	places where many people are. I don't *think* smoking should be completely banned at all the
file623697	people relax when they smoke. But I *think* smoking in a restaurant is bad manner. To enjoy
file623697	restaurant should be crime. I agree. I *think* smoking should be completely banned at all restaurants
file623697	And we want to enjoy eating. So I *think* smoking should be completely banned at all restaurants
file623697	banned at all restaurants in Japan. I *think* smoking should be completely banned at all the
file623697	smoking never has good effect on us. I *think* smoking should be completely banned at all the
file623697	customer who visit the restaurant. But I *think* smoking is very unkind thing for non-smoking people
file623697	disliked at almost places. However, I *think* smoking prohibited all the restaurants is not best
file623697	stimulates their appetite. And I *think* smoking at the table is very rude for the cook
file623697	all restaurants in Japan. " And I *think* smoking should be banned in other places. I don't
file623697	banned in other places. I don't *think* that smoking should be completely banned at all the
file623697	they can live more comfortable. I *think* smoking should be completely banned at all the
file623697	taste bad. Above two reasons make me *think* smoking should be completely banned at all the
file623697	country. There are three reasons why I *think* smoking should be completely controlled. First,
file623697	section and nonsmoking section so I *think* smoking don't need to be banned. Second, if smoking
file623697	dirty. By these reasons, I don't *think* smoking should be completely banned at all the
file623697	smoking makes delicious foods bad. I *think* smoking has no merit. But people who want to smoke
file623697	smoking. If they completely do that, I *think* smoking does not have to ban. Smoking itself is

Fig. 7 Edited concordance of *smoking* (n) (from the Japan ICNALE corpus)

The term *think* mostly occurs within the phrase *I think smoking*, thus inviting not only a first-person style which learners tend to like but also the corresponding formulaic phrase *should be (completely) banned/prohibited*. Of course, phrases such *as is very unkind thing* (Fig 7, Line 9) and *does not have to ban* (Fig 7, last line) indicate personal priming patterns that are out of sync with the majority of instances.

Turning now to the term *(part-time) job*, consider the following Word Sketches for Japanese, Hong Kong and Singaporean undergraduate/college students (see Figures 8, 9 and 10 respectively):



job *(noun)* icnalejapan-ptj freq = 2317 (23029.1 per million)

object_of	1556	2.8	subject_of	496	2.8	modifier	2449	2.8	predicate_of	111	5.8
have	1068	13.12	be	331	11.22	part-time	1576	13.62	chance	12	10.68
do	169	11.06	give	27	10.42	time	363	11.69	way	11	10.59
get	59	9.79	teach	14	9.59	part	260	11.58	step	6	10.35
experience	39	9.51	help	9	9.01	part-	90	10.18	need	5	10.3
think	40	9.43	make	11	8.66	many	16	7.46	practice	5	10.28
work	35	9.05	prevent	6	8.59	various	13	7.36	opportunity	6	10.14
start	19	8.61	tell	6	8.55	second	9	6.88	experience	17	9.89
choose	12	7.96	enable	5	8.35	full-time	8	6.74	place	6	9.83
find	11	7.81	have	26	8.21	other	6	6.24	thing	5	7.36
take	10	7.63	mean	4	7.99	good	6	6.06			

Fig. 8 Word Sketch for *(part-time) job* (n) (from the Japan ICNALE corpus)

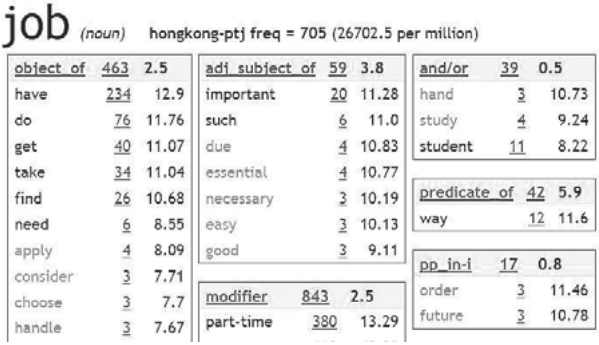


Fig. 9 Word Sketch for *(part-time) job* (n) (from the Hong Kong ICNALE corpus)

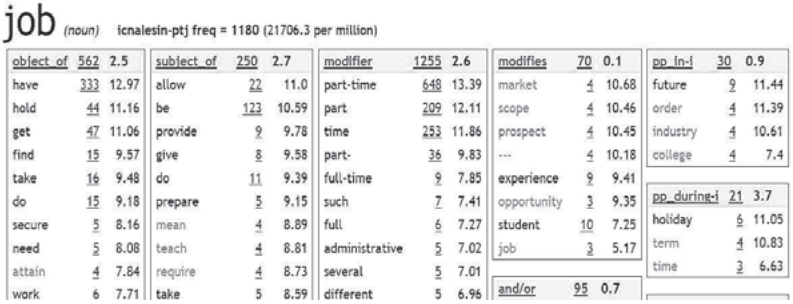


Fig. 10 Word Sketch for *(part-time) job* (n) from the Singapore ICNAL Corpus)

The verbs preceding the term *(part-time) job* for Figures 8, 9 and 10 are better compared with the figures generated for both the native learner (ENS1) and native speaker (ENS2) corpora in the ICNALE project (see Figures 11 and 12 respectively):

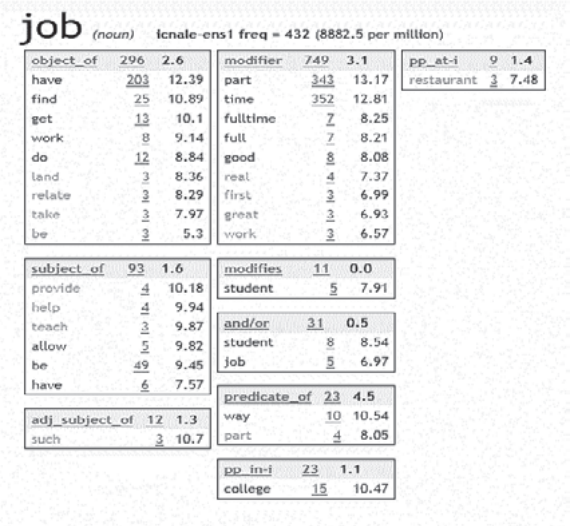


Fig 11. Word Sketch for *(part-time) job* (n) (from the ICNALE ENS1-native learner corpus)

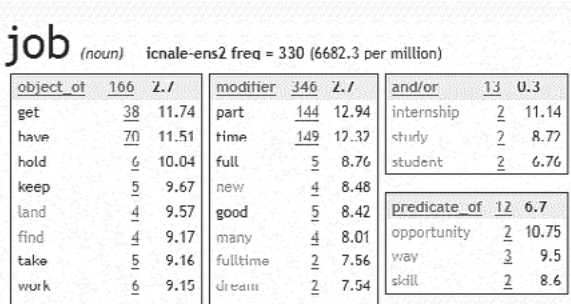


Fig 12. Word Sketch for *(part-time) job* (from the ICNALE ENS2-native speaker corpus)

VI Conclusion

The preceding sections have hopefully demonstrated the usefulness of the WordSketch software for automatically calculating and facilitating the frequencies of usage associated with (or 'prime') the top words *job* and *smoking* in the writing of college students in a number of the countries that have participated in the ICNALE project. Lexical priming is important for second language acquisition; global (native speaker) patterns of co-selection should be compared and contrasted with local (non-native speaker) ones in order to promote greater linguistic awareness and facilitate better

language learning.

References

Cheng, W. (2012). *Exploring corpus linguistics: language in action.* London: Routledge.

Flowerdew, J. (2009). Corpora in language teaching. In M Long & C Doughty (Eds.) *The Handbook of Language Teaching* (pp.327-350). West Sussex, UK: Wiley-Blackwell.

Hoey, M. (2005) *Lexical priming: a new theory of words and language*. London: Routledge.

Ishikawa, S. (2011). A new horizon in learner corpus studies: the aim of the ICNALE project. In G Weir, S Ishikawa, & K Poonpon (Eds.), *Corpora and language technologies in teaching, learning and research* (pp.3-11). Glasgow, UK: University of Strathclyde Press.

Kilgarriff, A. (2009) Corpora in the classroom without scaring the students. In *Proceedings of the 18th International Symposium on English Teaching*, Taipei. http://www.kilgarriff.co.uk/Publications/2009-K-ETA-Taiwan-scaring.doc [Accessed 2013-03-01]

Kilgarriff, A. & D. Tugwell (2002) Sketching words. In Marie-Hélène Corréard (Ed.) in *Lexicography and Natural Language Processing: A Festschrift in Honour of B. T. S. Atkins.* (pp. 125-137).
http://www.kilgarriff.co.uk/Publications/2002-KilgTugwell-AtkinsFest.pdf
[Accessed 2013-03-01]

Tarone, E., & B. Swierzbin (2009) *Exploring learner language*. Oxford: Oxford University Press.