Semiautomatic Extension of CoreNet using a Bootstrapping Mechanism on Corpus-based Co-occurrences

Chris BIEMANN
NLP Group, CS Institute,
University of Leipzig
Augustusplatz 10/11
04275 Leipzig, Germany
biem@informatik.uni-leipzig.de

Sa-Im SHIN
KORTERM, Division of
Computer Science, KAIST
373-1 Kusung Yusong
Daejon 305-701, Korea
miror@world.kaist.ac.kr

Key-Sun CHOI
KORTERM, Division of
Computer Science, KAIST
373-1 Kusung Yusong
Daejon 305-701, Korea
kschoi@cs.kaist.ac.kr

Abstract

The paper describes a language-independent approach for semiautomatic extension of lexical-semantic word nets and evaluates the method on CoreNet, the Korean version of word net. In a bootstrapping fashion, the so-called 'Pendulum Algorithm' operates on word sets obtained by co-occurrence statistics on a large un-annotated corpus and keeps error propagation low by a verification step. Results are not sufficient for automatic extension, but provide a good candidate set. Further improvements are discussed.

1 Introduction

A constantly addressed problem in computational linguistics and automated language processing is the so called 'acquisition bottleneck': A lot of time and money is being invested in building hand-crafted lexical resources for the use in further processing. Well-known resources in this respect are lexical-semantic word nets, such as WordNet (Miller, 1990), EuroWordNet (Bloksma et al., 1996) or CoreNet (Choi and Bae 2004). These word nets are widely accepted and used, despite their coverage problems: None of the nets contains significantly more than 100'000 lexemes of one language, whereas millions of lexemes can be found in corpora of decent size. Another problem is that the hierarchy in these nets is defined once and for all by the according linguists with dictionaries at hand and may or may not fit the domain in which it is going to be used. Even when using general-domain corpora like the Wall Street Journal, (Roark and Charniak, 1998) report that about 60% of the terms generated by their semantic class learner could not be found in WordNet. We believe that semi-automatic methods to find candidates for the extension of those nets will aid acquisition significantly and widen the bottleneck. In this paper, we present a bootstrapping approach operating on co-occurrence statistics of a large, unannotated corpus.

1.1 Related work

The use of bootstrapping approaches in order to assign words to semantic classes (which can be viewed as coarse-grained subtrees of WordNet-like hierarchies) has been gaining popularity in the recent years. Bootstrapping seems a viable way to obtain large amounts of data with merely a handful of start items (that can be rapidly prepared) by iteratively using all information previously learned in order to gain new information. The largest problem that bootstrapping methods have to face is error-propagation: misclassified items will acquire even more misclassified items. Various attempts have been made to minimize this thread.

(Riloff and Sherpherd, 1997) describe a method that assigns subject area categories to words using context similarity with pre-categorized words. However, error propagation is high and only 25% of the results were verified by human decision. In (Riloff and Jones, 1999) not only classes are assigned to words, but also the confidence of contexts supporting a class is estimated. Moreover, only the top 5 candidates are added to the knowledge per step, alleviating error-propagation to a precision of about 46%-76% after 50 iterations. Further improvement was gained in (Thelen and Riloff 2002), where multiple categories are learned at the same time to avoid too large single categories consisting of a mixture with several other categories. All these approaches use pattern-based extraction mechanisms for candidates that require at least some knowledge about the target language.

Perhaps the most similar approach to this proposal is (Roark and Charniak, 1998), using the log-likelihood measure for noun co-occurrences in order to find more words for the given categories. However, it requires full parsing and operates on rather small corpora. While the authors stress the value of using specific constructions for narrowing down the search space, we will merely use statistical measures without syntactically preprocessing the corpus, which keeps our method language-independent. Our central assumption is that a corpus large enough will produce enough senseful significant co-occurrences (see section 1.3) so that we can do without parsing.

1.2 CoreNet

CoreNet is an ontology containing three languages – Korean, Japanese and Chinese – constructed in KAIST[1]. The basic high-level concept of CoreNet is the "NTT basic concept hierarchy" constructed in Japan, and its structure lines up all semantic information.

CoreNet constitutes of 2,954 concepts that reflect Korean concepts. In this ontology, "concept" means a position in the semantic hierarchy and the term "sense" refers to the different meanings of a word form. Another important feature of CoreNet is that it uses the same concept hierarchy for nouns, verbs and adjectives. The major part of the vocabulary in northeastern languages – especially Chinese, Korean and Japanese – is derived from Chinese words and letters. For example, "N-hada" and "N+suru" are the Korean and Japanese version of a basic pattern "do+N" in English. In addition to the cultural sharing, this common feature in these languages makes it easy to combine them into the united ontology (see Choi et al. 2004). So, CoreNet is the overall ontology expressing Northeast-Asian language concepts.

The size of the CoreNet is as follows:

- Korean and Japanese: 28,823 nouns (56,523 senses), 1,757 verbs (4,717 senses), 804 adjectives (1,392 senses)
- Chinese: 20,647 nouns (28,932 senses), 288 verbs (765 senses), 80 adjectives (119 senses)

The size of Korean and Japanese is identical in CoreNet because of the constructing process of this ontology. The basic concept hierarchy for CoreNet is the "NTT basic concept hierarchy". We translated this Japanese hierarchy into Korean. After translating Japanese vocabulary in the NTT ontology, we mapped and adjusted the translated results based on this translated Korean hierarchy. In the case of Chinese vocabulary, we manually mapped it to CoreNet after finishing works on Korean and Japanese. We selected the Chinese vocabulary in CoreNet with the results of the accumulated information while making a Chinese-Korean dictionary for the translation system (see Zhang and Choi 1999).

The figure in Appendix A shows a screenshot of the Korean-Japanese noun hierarchy in CoreNet. The screen has four windows. The upper left side of the window shows a correspondence between Japanese and Korean words and concept numbers. The lower left side of the window contains word senses and definitions in the dictionary (Hangeul Society, ed. 1997). The upper right side of the window shows all words under a concept QUANTITY numbered 2588. The lower right side of the window shows a part of the list of concept hierarchy.

1.3 Statistically significant co-occurrences

Our major source for finding candidates is the notion of sentence-based statistical co-occurrence. The repeated occurrence of two or more words within a well-defined unit of information (sentence, document) is called a statistical co-occurrence. For the selection of meaningful and significant co-occurrences, an adequate co-occurence measure has to be defined. We use a significance measure similar to the well-known log-likelihood measure: Given two words *A*, *B*, each occurring *a*, *b* times in sentences, and *k* times together, we calculate the significance *sig(A, B)* of their co-occurrence in a sentence as follows:

$$sig(A,B) = x - k\log x + \log k!$$

$$\text{with } n = \text{number of sentences,}$$

$$x = \frac{ab}{n}.$$

Calculations are usually performed on very large corpora (>100 Million Tokens), using sentences or immediate neighboring words (sentence-based and neighborhood-based collocations, cf. Heyer et al. (2001)) as units. From an intuitive point of view, significant co-occurrences of a word w contain all kinds of associated words, be it typical modifiers, synonyms, antonyms, hyponyms, or members of the same semantic frame. Hence, the co-occurrence set of w contains words that are closely related to w. With the set of words comes a ranking for each word, based on the significance measure.

For the experiments in section 3 we used the Korean Version of Wortschatz (see http://www.wortschatz.uni-leipzig.de and (Biemann et al. 2004)). The number of sentences is

[1] http://bola.or.kr/eng/

about 2.3 Million, there are roughly 38 Million tokens and 3.8 Million distinct word types (inflected forms) of which over 0.9 Million have at least one significant co-occurrence, 430'000 words have at least five and 288'000 words have at least ten.

To give a short impression of how the significant co-occurrences look like, table 1 contains significant collocations for 연필(from the Korean corpus, meaning 'pencil') and jurisdiction (from our English corpus), together with their significance values.

reference word	**TOP 25 co-occurrences ordered by significance**
연필 (pencil)	지우개 (eraser) (25), 만년필 (fountain pen) (22), 국어 (Korean) (14), 볼펜 (ball pen) (14), 쥐는 (grasping) (14), 한자루도 (a pen) (14), 한쪼가리 (a part of) (14), 문구세트 (stationary set) (13), 문화연필은 (Mun-Hwa pencil) (13), 자루 (the measure of numbering pencils) (11), 필통 (pencil box) (11), 한토막 (a part) (11), 공책 (notebook) (10), 기념품을 (souvenir) (9), 노트 (notebook) (9), 시간 (time) (9), 그린 (drawing) (8), 사진 (picture) (8), 한글을 (Korean) (8), 가방 (bag) (7), 쓰던 (writing) (7), 쓰면 (writing) (7), 아이들은 (children) (7), 종이 (paper) (7), 줄은(decreasing) (7) [..]
jurisdiction	over (305), court (188), under (183), courts (145), federal (121), Court (95), case (73), court's (68), state (45), within (43), Appeals (38), ruled (38), Circuit (36), SEC (36), law (36), Commission (34), GSBCA (34), appeals (34), House (33), committees (33), Judge (31), Act (29), CFTC (29), Committee (29), subcommittee (28) [...]

Table 1: Examples of significant co-occurrences

2 The Pendulum Algorithm on Co-occurrence sets

Looking at significant co-occurrences, the following observations can be made: first of all, the co-occurrence sets include many words that are closely related to the reference word in a CoreNet or WordNet sense. Second, there are many words that are not wanted in WordNet close to the reference word: some are stop words, some belong to unwanted part-of-speeches and some are not paradigmatically related at all. Hence, co-occurrence sets itself are not pure enough for enhancing lexical-semantic word nets or semantic categories, but they can serve as a data basis where our algorithm works upon.

The pendulum algorithm was first described in (Quasthoff et al. 2002), where it was used for the detection of person names in large corpora using pattern rules on flat word features and a small seed set of 19 name parts, blowing up the number of learnt name parts with a factor of over 2000 without considerably losing on precision. Its power lies in the alternation of a search step, where candidates for knowledge extension are determined, and a verification step in which the candidates are checked and accepted or rejected.

The algorithm is reformulated for co-occurrence sets as follows: For each item in the (initially small) knowledge, get the co-occurrence set. Elements of the co-occurrence set are candidates for extension. For all the candidates, obtain the co-occurrence sets and check how many words in the co-occurrence set of each candidate are already in the knowledge. If this number is above a certain threshold, accept the candidate and reuse it later in the search step. Figure 1 states the algorithm in pseudo-code.

```
LastLearned=StartSet;
Knowledge=StartSet;
NewLearned=0;
while (LastLearned>0) {
  for all i in LastLearned {
  Candidates=getCooccurrences(i);
    for all c in Candidates {
      VerifySet=getCooccurrences(c);
      if |VerifySet ∩ Knowledge|
         >threshhold {
        NewLearned+=c;
        Knowledge+=c;
      }
    }
  }
  LastLearned=NewLearned;
  NewLearned=0;
}
```

Figure 1: the pendulum algorithm on co-occurrence sets

Parameters of the algorithm like threshold and the number of co-occurrences used in the search and in the verification step crucially rely on corpus size and the size of the start set. When using small corpora, there will be only small sets, resulting in a smaller threshold. This, however, reduces precision.

To rule out too common words (stop-words), the `getCooccurrences()`-function only returns words below a certain corpus frequency. The words filtered out by this mechanism would infect the knowledge soon and result in very poor learnt knowledge sets of general nature, instead of the specialized knowledge sets we went to obtain here. Further, parameters can be set in order to reduce the size of the co-occurrence sets by cutting of low significant elements.

To a deeper understanding of the process, let us closely examine an example from CoreNet. English translations can be found in brackets behind the words.

We start with some words from concept number 555, related to head and face:

관자놀이 (temple), 눈 (eye), 빰 (cheek), 시(poem), 쌍꺼풀 (double eyelid), 아랫입술 (lower lip), 오관 (the five sensory organs), 입 (mouth), 코 (nose), 혀 (tongue),

The search step for "관자놀이" (temple) returns the single candidate "복사뼈" (malleolus bone), which verifies through the three words in its co-occurrence set marked in bold:

부위마다 (part of face), 안면부 (part of the face), 인당 (ligament), 인중 (philtrum), 경골 (tibial), 관자놀이 (temple), 경혈을 (spots on the body suitable for acupuncture), 손끝으로 (with fingertip), 용천 (spring), 청명 (serenity), 4차례씩 (per 4 times), 두드릴 (tabbing), 발바닥 (the sole of the foot), 코와 (with nose), 등 (back), 오리 (duck), 상부 (high part), 위쪽 (front part), 신체 (body), 예방하는 (preparing), **입 (mouth)**, 질병을 (disease), **코 (nose)**, 한가운데 (center), 가볍게 (lightly), 곳 (place), 누르고 (pressing), 영향 (influence), 중간 (middle), 지정된 (appointed).

Another search step for "입" returns amongst others the candidates 입술(mouth) and 눈먼(becoming blind). 입술 verifies through 입, 빰 and 혀 (tongue) whereas "눈먼" is rejected because of no support.

The ambiguous word "분" (minute/make-up powder) is not verified because its co-occurrence set only contains two words: 시 and 입. The last case - a possible unit term - would be a good source for infection of the word set.

The previously learned item 입술(mouth) finds the accepted candidate 눈썹 (eyebrow).

Usually, far more candidates are rejected then accepted due to the careful verification mechanism which trades recall for high precision, which is the main parameter to optimize in bootstrapping - rejected candidates in early steps can be accepted later when more items have been learned already.

3 Evaluation

3.1 Methodology

In order to evaluate this method, we applied the pendulum algorithm in order to extend the CoreNet coverage. Co-occurrence statistics where obtained from an analysis of the KAIST corpus[2]. KAIST corpus is a large-scaled Korean raw corpus, which contains about 40 million eojeol in 2355860 sentences. For the experiments, we used sentence-based co-occurrences.

We selected 15 subtrees of the CoreNet hierarchy of different sizes at random. For the search step and for the verification step, we used the most significant 100 co-occurrences (if available), ordered by significance.

The verification threshold was varied as follows: going down from a support of eight to three co-occurrences in the verification step, we evaluated the result for the highest threshold that did not produce a result set larger than the start set. This heuristics was applied in order to detect and avoid result set infection. To exclude high-frequency words from closed word classes we rejected the 1000 most frequent words in the corpus.

To avoid biased results due to ad-hoc parameter tuning and circumnavigating bad examples, a non-Korean speaker performed the selection of concepts.

In the evaluation, only words that have not been mentioned in the CoreNet subtree before were taken into account. Table 2 shows the characteristics of the different start sets.

CoreNet ID	Name of Concept	Number of Members
50	human good/bad	119
51	baby, children	43
111	human relation	274
113	partner / co-worker	123
114	partner / member	71
181	human ability	213
430	store	128
471	land, area	260
548	insect, bug	75
552	part of animal	736
553	head	139
577	forehead	72
590	legs and arms	86
672	plant (vegetation)	461
817	cloths	246

Table 2: Evaluated concepts

[2] http://kibs.kaist.ac.kr

3.2 Evaluation results

Evaluation of the results had to be done manually, because we only counted new words. A typical run needed 3-7 iterations to converge. Table 3 shows the precision of the algorithm for the different concepts.

CoreNet ID	# new words	# correct	precision
50	36	5	13.89%
111	3	2	66.67%
113	23	8	34.78%
114	5	3	60.00%
181	7	2	28.57%
430	12	11	91.67%
471	10	2	20.00%
548	43	6	13.95%
552	10	6	60.00%
553	7	4	57.14%
577	4	2	50.00%
590	7	3	42.86%
672	30	15	50.00%
817	34	18	52.94%
Sum	**231**	**87**	**37.67%**

Table 3: Evaluation results by concept

Note that the size of the start set does not correlate with the numbers of new words. This is due to the fact that some concepts have a higher frequency than others in the KAIST corpus, which mainly consists of law, government and economy texts.

Common errors consisted of functional words and words that are likely to appear in the same context as the words contained in the concept, but do not generally belong to that concept.

While results are far from being useable for fully automatic extension of the CoreNet resource, they still provide a viable way of finding candidates for extension by simply feeding an unannotated corpus into the sequential process of statistical analysis and the pendulum algorithm.

Note that this process is completely language independent in a way that neither the co-occurrence analysis nor the pendulum algorithm makes any assumption about the language of the corpus. The results presented here may serve as baseline for what is reachable for every language, given concepts containing at least 50 words.

In the remainder, we will discuss how further improvement may be gained by dropping the claim on language-independence.

4 Further Work

To avoid the problem of having stop words like functional words in the result sets it is possible to use a POS-tagged corpus and to look only for the word classes of interest, i.e. verbs, adjectives and nouns. Moreover, some experiments with German language (see Biemann et al. 2004) showed improvement when using different word classes for search and verification. A more thorough evaluation is needed to confirm this, however.

Another issue is recall: while the KAIST corpus contains about 3,8 Million word forms, less then 10% of them have more then 10 statistically significant co-occurrences. Therefore, many correct candidates for CoreNet extension cannot be found by the proposed method. There are two ways to alleviate the problem: one possibility is the use of a larger corpus. Another possibility would be to use more sophisticated methods to extract candidates, e.g. parsing (see Roark and Charniak 1998). This, in turn, requires much more manual work for building tree banks and training parsers in advance.

The evaluation of methods described in (Biemann et al. 2004) that try to determine the appropriate relation between words in order to automatically construct WordNet-like structures will be in focus of our follow-up research.

References

C. Biemann, S. Bordag, G. Heyer, U. Quasthoff and C. Wolff. 2004. *Language-independent Methods for Compiling Monolingual Lexical Data*. Proceedings of CiCling 2004, Seoul, Korea and Springer LNCS 2945, pp. 215-228, Springer Verlag Berlin Heidelberg

L. Bloksma, Diez-Orzas and P. Vossen 1996. *User requirements and functional specification of the EuroWordNet project*. Deliverable D001, EuroWordNet, LE2-4003, Computer Centrum Letteren, University of Amsterdam. Amsterdam.

K.-S. Choi, Hee-Sook Bae. 2004. *Procedures and Problems in Korean-Chinese-Japanese Wordnet with Shared Semantic Hierarchy*, Global WordNet Conference, Czech Republic

G. A. Miller. 1990. *Wordnet - an on-line lexical database*. International Journal of Lexicography 3(4):235-312

U. Quasthoff, C. Biemann, and C. Wolff. 2002. *Named entity learning and verification: Expectation Maximisation in large corpora*. In: Proceedings of CoNLL-2002, The Sixth Workshop on Computational Language Learning, 31 August and 1 September 2002 in association with Coling 2002 in Taipei, Taiwan

E. Riloff and R. Jones. 1999. *Learning dictionaries for information extraction by multi-level bootstrapping*. In Proceedings of AAAI-99.

E. Riloff and J. Sherpherd. 1997. *A corpus-based approach for building semantic lexicons*. In Proceedings of the Second Conference on Empirical Methods in Natural Language Processing.

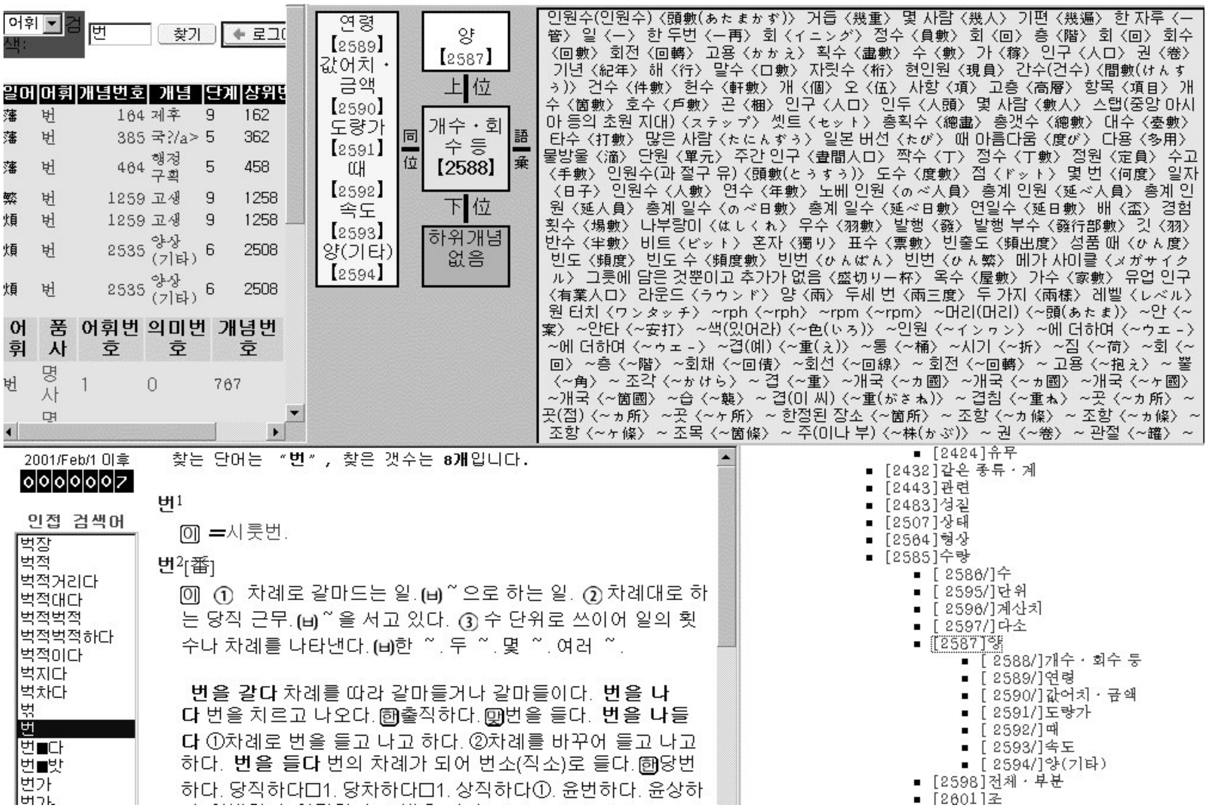
B. Roark and E. Charniak. 1998. *Noun-phrase Co-occurrence Statistics for Semi-automatic Semantic Lexicon Construction*. In Proceedings of the 36th Annual Meeting of the Association for Computantional Linguistics, pp. 1110-1116.

M. Thelen and E. Riloff. 2002. *A bootstrapping method for learning semantic lexicons using extraction pattern contexts*. In Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP 2002).

M. Zhang, K.-S. Choi 1999. *MATES/CK A Chinese-to-Korean Machine Translator*. In Proceedings of the 5th Natural Language Processing Pacific Rim Symposium (NLPRS'99), pp. 245-250, Beijing, China, November 1999

Hangeul Society, ed. .1997. *Urimal Korean Unabridged Dictionary*, Eomungag

Appendix A: Web browser for CoreNet



일어	어휘	개념번호	개념	단계	상위번
藩	번	164	제후	9	162
藩	번	385	국?/a>	5	362
藩	번	464	행정 구획	5	458
繁	번	1259	고생	9	1258
煩	번	1259	고생	9	1258
煩	번	2535	양상(기타)	6	2508
煩	번	2535	양상(기타)	6	2508

어휘	품사	어휘번호	의미번호	개념번호
번	명사	1	0	767
	명			

인원수(인원수) 〈頭數(あたまかず)〉 거듭 〈幾重〉 몇 사람 〈幾人〉 기편 〈幾遍〉 한 자루 〈一
管〉 일 〈一〉 한 두번 〈一再〉 회 〈イニング〉 정수 〈員數〉 회 〈回〉 층 〈階〉 회 〈回〉 회수
〈回數〉 회전 〈回轉〉 고용 〈かかえ〉 획수 〈畫數〉 수 〈數〉 가 〈稼〉 인구 〈人口〉 권 〈巻〉
기년 〈紀年〉 해 〈行〉 말수 〈口數〉 자릿수 〈桁〉 현인원 〈現員〉 간수(건수) 〈間數(けんす
う)〉 건수 〈件數〉 헌수 〈軒數〉 개 〈個〉 오 〈伍〉 사항 〈項〉 고층 〈高層〉 항목 〈項目〉 개
수 〈箇數〉 호수 〈戶數〉 곤 〈梱〉 인구 〈人口〉 인두 〈人頭〉 몇 사람 〈數人〉 스탭(중앙 아시
아 등의 초원 지대) 〈ステップ〉 셋트 〈セット〉 총획수 〈總畫〉 총갯수 〈總數〉 대수 〈臺數〉
타수 〈打數〉 많은 사람 〈たにんずう〉 일본 버선 〈たび〉 때 아름다움 〈度び〉 다용 〈多用〉
물방울 〈滴〉 단원 〈單元〉 주간 인구 〈晝間人口〉 짝수 〈丁〉 정수 〈丁數〉 정원 〈定員〉 수고
〈手數〉 인원수(과 절구 유) 〈頭數(とうすう)〉 도수 〈度數〉 점 〈ドット〉 몇 번 〈何度〉 일자
〈日子〉 인원수 〈人數〉 연수 〈年數〉 노베 인원 〈のべ人員〉 총계 인원 〈延べ人員〉 총계 인
원 〈延人員〉 총계 일수 〈のべ日數〉 총계 일수 〈延べ日數〉 연일수 〈延日數〉 배 〈盃〉 경험
횟수 〈場數〉 나부랑이 〈はしくれ〉 우수 〈羽數〉 발행 〈發〉 발행 부수 〈發行部數〉 깃 〈羽〉
반수 〈半數〉 비트 〈ビット〉 혼자 〈獨り〉 표수 〈票數〉 빈출도 〈頻出度〉 성품 때 〈ひん度〉
빈도 〈頻度〉 빈도 수 〈頻度數〉 빈번 〈ひんぱん〉 빈번 〈ひん繁〉 메가 사이클 〈メガサイク
ル〉 그릇에 담은 것뿐이고 추가가 없음 〈盛切り一杯〉 옥수 〈屋數〉 가수 〈家數〉 유업 인구
〈有業人口〉 라운드 〈ラウンド〉 양 〈兩〉 두세 번 〈兩三度〉 두 가지 〈兩樣〉 레벨 〈レベル〉
원 터치 〈ワンタッチ〉 ~rph 〈~rph〉 ~rpm 〈~rpm〉 ~머리(머리) 〈~頭(あたま)〉 ~안 〈~
案〉 ~안타 〈~安打〉 ~색(있어라) 〈~色(いろ)〉 ~인원 〈~インワン〉 ~에 더하여 〈~ウエ-〉
~에 더하여 〈~ゥエ-〉 ~겹(예) 〈~重(え)〉 ~통 〈~桶〉 ~시기 〈~折〉 ~짐 〈~荷〉 ~회 〈~
回〉 ~층 〈~階〉 ~회채 〈~回債〉 ~회선 〈~回線〉 ~ 회전 〈~回轉〉 ~ 고용 〈~抱え〉 ~ 뿔
〈~角〉 ~ 조각 〈~かけら〉 ~ 겹 〈~重〉 ~개국 〈~カ國〉 ~개국 〈~ヵ國〉 ~개국 〈~ヶ國〉
~개국 〈~箇國〉 ~습 〈~襲〉 ~ 겹(이 씨) 〈~重(がさね)〉 ~ 겹침 〈~重ね〉 ~곳 〈~カ所〉 ~
곳(점) 〈~ヵ所〉 ~곳 〈~ヶ所〉 ~ 한정된 장소 〈~箇所〉 ~ 조항 〈~カ條〉 ~ 조항 〈~ヵ條〉 ~
조항 〈~ヶ條〉 ~ 조목 〈~箇條〉 ~ 주(이나 부) 〈~株(かぶ)〉 ~ 권 〈~巻〉 ~ 관절 〈~躍〉 ~

2001/Feb/1 이후
0000007

인접 검색어

벅장
벅적
벅적거리다
벅적대다
벅적벅적
벅적벅적하다
벅적이다
벅지다
벅차다
벆
번
번■다
번■밧
번가
번가

찾는 단어는 "**번**", 찾은 갯수는 **8개**입니다.

번[1]

[이] **=**시룻번.

번[2][番]

[이] ① 차례로 갈마드는 일. (ㅂ)~으로 하는 일. ② 차례대로 하는 당직 근무. (ㅂ)~을 서고 있다. ③ 수 단위로 쓰이어 일의 횟수나 차례를 나타낸다. (ㅂ)한 ~. 두 ~. 몇 ~. 여러 ~.

번을 갈다 차례를 따라 갈마들거나 갈마들이다. **번을 나다** 번을 치르고 나오다. [한]출직하다. [맞]번을 들다. **번을 나들다** ①차례로 번을 들고 나고 하다. ②차례를 바꾸어 들고 나고 하다. **번을 들다** 번의 차례가 되어 번소(직소)로 들다. [한]당번하다. 당직하다□1. 당차하다□1. 상직하다①. 윤번하다. 윤상하

- [2424]유무
- [2432]같은 종류 · 계
- [2443]관련
- [2483]성질
- [2507]상태
- [2564]형상
- [2585]수량
    - [2586/]수
    - [2595/]단위
    - [2596/]계산치
    - [2597/]다소
    - [2587]양
        - [2588/]개수 · 회수 등
        - [2589/]연령
        - [2590/]값어치 · 금액
        - [2591/]도량가
        - [2592/]때
        - [2593/]속도
        - [2594/]양(기타)
    - [2598]전체 · 부분
    - [2601]조