Automatic extraction of subcategorization frames for French

Paula Chesley[*], **Susanne Salmon-Alt**[†]

[*]Linguistics Department, University at Buffalo, USA
pchesley@buffalo.edu
[†]ATILF-CNRS, Nancy, France
salt@atilf.fr

Abstract

This paper describes the automatic extraction of French subcategorization frames from corpora. The subcategorization frames have been acquired via VISL, a dependency-based parser (Bick, 2003), whose verb lexicon is currently incomplete with respect to subcategorization frames. Therefore, we have implemented binomial hypothesis testing as a post-parsing filtering step. On a test set of 104 frequent verbs we achieve lower bounds on type precision at 86.8% and on token recall at 54.3%. These results show that, contra (Korhonen et al., 2000), binomial hypothesis testing can be robust for determining subcategorization frames given corpus data. Additionally, we estimate that our extracted subcategorization frames account for 85.4% of all frames in French corpora. We conclude that using a language resource, such as the VISL parser, with a currently unevaluated (and potentially high) error rate can yield robust results in conjunction with probabilistic filtering of the resource output.

1. Introduction

The *subcategorization frame* (SCF) of a verb specifies the number and categories of syntactic arguments a verb takes. As such, SCFs constitute a well-studied linguistic phenomenon from a theoretical perspective. SCFs are also of immediate utility in natural language processing (NLP) for electronic dictionaries and statistical parsers.

(Briscoe and Carroll, 1993) observe that half of all automatic parse errors for English stem from incorrect subcategorization frames. Inaccurate prepositional attachment is an example of this type of error. A parser with access to subcategorization information can make a better informed choice about whether to attach the prepositional phrase (PP) in (1) correctly to the verb phrase (VP), and not to the direct object noun phrase (NP).

(1) a. * Sarah [put [the ball [on the table]]].

b. Sarah [put [the ball] [on the table]].

SCFs can help to override parsing heuristics such as structural biases for attaching the PP to the nearest dominating node, thereby improving higher-level attachment accuracy.

Despite these claims, state-of-the-art lexicalized parsers for English and French have shown only modest improvements when subcategorization information is added. (Collins, 1999) integrates subcategorization information into a head-driven lexicalized parser for English. His model 2 including subcategorization information improves F-measure over model 1, with no subcategorization information, by .8%. This relatively low increase is due to the considerable overlap in subcategorization information with the distance measure employed. But when the distance measure includes neither adjacency nor verb conditions, subcategorization yields a 10% improvement in parser accuracy. Collins remarks that for English, if one had to choose between including distance or subcategorization, distance would be preferable from an engineering perspective. What's more, (Arun, 2004) shows that when parsing French, subcategorization information, viz., an emulation of Collins' model 2 (Collins, 1999), shows only a statistically insignificant increase over the French model 1.

Yet we should not exclude the potential utility of subcategorization information when parsing French. First, as Arun notes, the insignificant results are partially due to the flat annotation scheme of the French Treebank (Abeillé et al., 2003), which does not lend itself well to identifying SCFs. That is, the treebank structure could prevent the increases we see in Collins' results. Second, since (Arun, 2004) does not include a detailed error analysis with respect to SCFs, we cannot quantify to what extent subcategorization information is useful *on specific structures*. Collins acknowledges that choosing distance over subcategorization information may not be preferable for languages with freer word order, especially those in which complements can appear to the right or left of the head. We suspect that a French parser could benefit from subcategorization information on many constructions with non-canonical constituent order, like topicalization, an example of which is given in (2):

(2) Le vin rouge j'aime bien mais je préfère le rosé.
The wine red I-like well but I prefer the rosé
"Red wine I like but I prefer rosé"

When parsing oral French, it is highly desirable to have a mechanism to account for non-canonical realizations like topicalization. Additionally it is not clear how a distance-only parser would fare on ambiguous pre-verbal object clitics, such as *vous*, "you", which can be either a direct or indirect object, or the causative construction, in which arguments are either non-canonically realized or optional. Finally, although lexicalized statistical parsers have shown great success in recent parsing endeavors, it is entirely possible that subcategorization information could be of great use in other parsing formalisms to which not as much research has currently been devoted.

The current work distinguishes itself from other work on automatically extracting SCFs from corpora in several respects. Crucially, the immediate goal of this work is application-independent: our results will not be used in

conjunction with solely one parser, but rather, will be annotated in the Lexical Markup Framework, an emerging International Standards Organization annotation standard (ISO TC 37/SC 4) for lexical databases (Romary et al., 2004). This work, as that of (Sarkar and Zeman, 2000), is one of the few in which the frames are not determined a priori. A benefit of this approach is that one is not limited to a small number of pre-determined frames if an extensive subcategorization frame lexicon does not exist for the language of interest (or if the researcher does not want to create a small subcategorization lexicon ad hoc). We also estimate the coverage and generalizability of our results and interpret our findings in the wake of recent methodological issues raised about binomial hypothesis testing.

2. Method

2.1. Corpus

We created a multi-genre corpus of 40-125 random occurrences of 104 frequent verbs using the Frantext online literary database[1]. The Frantext resource is available at `http://www.frantext.fr/categ.htm`. Since it has date- and genre-delimiting fields, we limited the dates to between 1850 and 2000 so as not to obtain any archaic SCFs. We also exluded the poetry and theatre genres so as to minimize potential SCF noise these genres might produce. We queried the tagged version of this database, taking advantage of the Frantext query language to eliminate the maximum number of noisy causative and *ne...que*, "only", constructions. The former construction permits non-canonical or optional argument realization, and the latter includes *que*, a function word that is notoriously hard to tag and parse in French.

2.2. Pre-filtering

Eckhard Bick then parsed this corpus with the VISL parser (Bick, 2003). This parser is not yet fully lexicalized, but it does have a subcategorization lexicon for some verbs. We chose to examine 80 random verbs in this lexicon. We selected 24 other verbs either because they were in the Test Suites for Natural Language Processing for French (Estival and Lehmann, 1997), or were otherwise highly frequent in our sample. Verbs in the Test Suites are considered a balanced sample of the French verbal lexicon. Wanting to see if our results are dependent on whether the VISL subcategorization lexicon includes the verbs in our sample, we also included other frequent verbs not in the VISL lexicon.

Since we do not assume a set of a priori frames to be extracted, the frames are determined throughout the filtering stage. However, we limit the syntactic constituents we count as possible SCF elements to the following, conflating syntactic categories and syntactic functions for ease of exposition:

- direct objects;
- PPs headed by a particular preposition. Indirect objects are subsumed under the PP headed by *à*, "to";
- subordinate clauses and small clauses with various heads;
- infinitive verbs, in the case of raising and control verbs;
- predicative adjectival phrases;
- reflexive clitic NPs.

The resulting SCFs can consist of any combination of the above elements. We do not include subjects in our discussion of SCFs or SCF constituents since French verbs must take a grammatical subject.

Elements in the above list were chosen for various theoretical and practical concerns. Exactly what constitutes a subcategorization frame is a debatable matter, and one which, for reasons of space and scope, we do not discuss in the current work. Rather, combinations of the above elements are the most productive in French, and since they are also the most frequently occuring constituents in French corpora, we can actually interpret the corpus evidence for or against them in terms of SCFs. We count reflexive clitics as subcategorizable constituents, despite the fact that for some verbs, reflexive clitics are arguably not regarded as part of the lemma (e.g., *se laisser VINF* and *laisser quelqu'un/quelque chose VINF*, "to let oneself VINF" and "to let someone/something VINF", respectively). Some reflexive clitics are indeed part of the lemma (e.g. *se pencher sur quelque chose*, "to look into something", vs. *pencher quelque chose, e.g., un regard, sur quelqu'un*, "to dart something, e.g., a look, at someone"), and there is no immediately obvious way to distinguish in a syntactically unannotated corpus these two uses of reflexive clitics.

Upon examining the parser output, we saw that the parser conflates modifiers and complements. Hence we decided to strip all modifier/complement distinctions, making anything that was a sister of the verb in the parser output a constituent for which the verb could potentially subcategorize. This approach assumes the filtering stage will eliminate the incorrect frames. Also, it generalizes well to verbs not included in the VISL subcategorization lexicon. It is worth noting that with this approach many true SCFs will be embedded in other erroneous frames proposed by the parser.

2.3. Filtering

We implemented binomial hypothesis testing (BHT) to filter the noisy parser output. In our implementation of BHT, the erroneous subsuming frames discussed in the previous section neither count as evidence for nor against observing a true frame embedded inside them.

BHT in this application examines the difference between the number of times a particular cue occurs with a given verb and the number of total times the latter appears in the corpus, where a *cue* is an initial frame we receive from the parser, without knowing whether it is indeed a frame for the given verb. The greater this difference, the less likely it is that the cue is an actual frame. Let m be the total number of occurrences of a verb in the corpus, n be the number of co-occurrences of the verb with the cue, and B_f the estimated probability that the verb that does not subcategorize for the frame f appears nevertheless with f. We make the null

[1] A previous work (Chesley and Salmon-Alt, 2005) examines 200 occurrences of 115 verbs. These numbers do not take into account the large data loss we incurred upon transforming the VISL output to an exploitable, interpretable format.

hypothesis that the verb does not subcategorize for the cue. The upper bound on the probability that the hypothesis is false given all cues is the following (Manning, 1993):

$$P(n+, m, B_f) = \sum_{i=n}^{m} \frac{m!}{i!(m-i)!} {B_f}^i (1 - B_f)^{m-i}$$

In the present work we have set the confidence level at .02, below which the cue is considered an actual frame. Typical confidence levels are empirically set between .02 and .05, according to user needs and sample size. Section 4. deals with the issues of confidence levels and sample size in greater detail.

We adapted the method used in (Brent, 1993) to find the error rate B_f for each frame. Brent's original method consists in examining the first x occurrences of a frame with every verb in the corpus above x occurrences, where x in the current work is fixed at 40. From these occurrences we construct a histogram based on the number of co-occurrences of a frame and the verbs with a sufficient amount of corpus attestations. An example of such a histogram, akin to Brent's figure 1, is given in figure 1. At the lower end of the histogram we expect to see binomially distributed (i.e., Gaussian-like) noise, a pattern that represents the verbs that do not subcategorize for the frame, but that appear with it nevertheless. Beyond this noise we assume that the co-occurrence of a verb with the frame is not random. The mean co-occurrence probability in the binomial distribution at the low end of the histogram is therefore a proper estimation of the rate of false frames B_f.

In order to determine the mean B_f, we first estimate a cutoff point, the end of the right tail of the binomial distribution, under which all verbs have similarly low co-occurrence frequencies with the frame. For example, the cutoff point in figure 1 is 10. Above this cutoff point, the verbs have higher, less normally distributed co-occurrence frequencies with the frame. The estimation of this cutoff point is an iterative procedure in which first a bin from the histogram is tried as an approximation of the cutoff point. Second, with this cutoff point, we estimate the mean of the binomial distribution, which leads to an expected shape of the distribution. We then compare the expected shape of the distribution with its actual shape, and choose the predicted mean value that minimizes the squared error of the binomial distribution. B_f is then the maximum-likelihood probability of co-occurrence of the verbs and the frame at this mean.

If the frames are not known a priori, it is difficult to see how Brent's method could be effective if not modified slightly. Specifically, this method requires a significant number of identical frames in order to form a binomial distribution at the low end of a histogram like that in figure 1. In our data, the frames output by the parser include both modifiers and complements. Identical frames are relatively scarce, and we would rarely see the distributional regularities in the data required by this method. However, each subcategorized constituent in the frame does appear frequently. To render the problem tractable given our data, we make the additive independence assumption, whereby the error rate of an entire frame is the sum of all potentially subcategorized constituents in the frame. We thus determine the error rate for each constituent in a frame and sum over all constituents in the frame to find the error rate B_f for the frame.

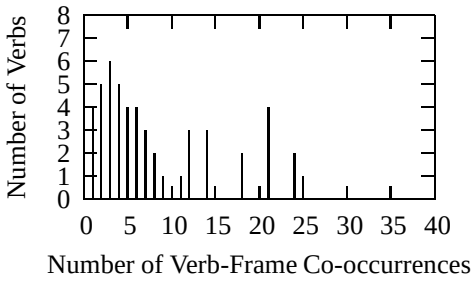

Figure 1: A histogram to be used for automatically determining the error rate of a frame, akin to figure 1 of (Brent 1993). The binomially distributed noise at the low end of the histogram shows the verbs that do not subcategorize for the frame.

Two points about our adaptation of Brent's method are worth noting. The first is that our implementation has a bias toward frames of shorter length. The lower the error rate B_f, the more likely a proposed frame is to be an actual frame. Since we sum over all constituents, it is natural that SCFs with a greater number of constituents will have a larger error rate. To some extent this bias is overly simplistic; nevertheless, it does correspond to the general trend that the lower the number of constituents a subcategorization frame has, the more probable it is in a corpus. Second, the independence assumption we make of constituents in effect says that we examine only marginal probabilities of constituents, and not possible interactions between constituents, where indeed such interactions may exist. Nevertheless, the independence assumption has been shown to be robust to minor violations in many NLP applications.

3. Results

Our results from this experiment are 27 unique SCFs and 176 verb-frame combinations. The frequencies of our unique frames are given in table 1. To evaluate this work, we submitted our results to two native speakers of French who are also computational linguists. This manual method of evaluation was chosen since currently there is no electronic resource explicitly encoding SCFs for French. The criteria given to the evaluators take into account the level of familiarity the evaluators have with the subject and account for both syntactic and semantic verb behavior.

An evaluation of SCFs on both syntactic and semantic levels is necessary for several reasons. First, only 2% of French verbs take obligatory complements (Guillet and Leclère, 1992). This low figure quantifies our intuition that given an appropriate context, most semantically obligatory arguments can be interpreted implicitly. Given the optional status of many arguments, syntactic criteria alone yield an incomplete evaluation of SCFs. Hence the essence of our evaluation method is to determine whether the SCF (1) expresses a semantic argument of the verb, as opposed to an adjunct, (2) uses the proper syntac-

Frame	Frequency
V DO	62
V (intrans.)	28
V VFIN$_{que}$	19
REFL V	15
V VINF$_{de}$	6
V attributive adj.	4
V VINF$_{\grave{a}}$	4
REFL V PP$_{\grave{a}}$	4
V PP$_{par}$	4
REFL V VINF$_{\grave{a}}$	3
V PP$_{\grave{a}}$	3
V DO PP$_{\grave{a}}$	3
V VINF	3
V DO PP$_{de}$	2
V PP$_{de}$	2
REFL V PP$_{de}$	2
REFL V PP$_{sur}$	2
V VINF$_{par}$	1
REFL V PP$_{vers}$	1
REFL V VINF	1
REFL V PP$_{par}$	1
REFL V VINF$_{de}$	1
REFL V PP$_{en}$	1
REFL V VINF$_{\grave{a}}$	1
V PP$_{dans}$	1
V DO VINF	1
V PP$_{\grave{a}}$ VINF$_{de}$	1
Total	176

Table 1: SCFs our system identifies, along with their frequencies in our sample. V = verbal entry, REFL = reflexive clitic, PP$_x$ = PP headed by x, and VFIN$_x$, VINF$_x$ are respectively finite and non-finite clauses, headed by x. Note the classic Zipfian distribution of the results.

tic constituents in expressing the semantic argument, (3) is seen in the *Trésor de la Langue Française informatisé*, an exhaustive French dictionary, if one is unsure about the proposed frame.

Inter-rater agreement was judged reliable, at $K = .82$. We take the lower bound for precision to be the intersection of SCFs both raters judged correct; this figure is 86.8%. Conversely, the upper bound for precision, 96.4%, is the union of the frames at least one rater judged felicitous. These figures do not take into account six SCFs proposed by our system that had subcategorized constituents other than those we intially sought as subcategorizable; five of these were the impersonal subject *il*. Determining an impersonal subject seemed beyond the scope of our method of extracting SCFs, so we leave these errors for subsequent research. The baseline for this task, simply guessing the most common SCF, direct object, would yield an F-measure of 35.2%.

Token recall was found to be 54.3%. This figure was arrived at by examining four random occurrences of each verb from a corpus of online French newspaper articles. Comparatively, the token recall rate of (Manning, 1993), calculated similarly to ours, is 82%. However, Manning's results show a similar ratio of learned frames to verbs to ours (1.58 vs. 1.69, respectively). We had initially posited that the high frequency and polysemy of our sample verbs were causing this disparity in results, since Manning's recall is for random verb occurrences. Since the 104 verbs in our sample are extremely frequent and polysemous, it is natural to assume that they exhibit many different SCFs (if we assume the number of SCFs to be directly related to the number of verb senses). Yet we did not find the expected inverse relationship between number of senses and our recall rates. A preliminary investigation shows that frame recall is inversely related simply to the number of frames a verb accepts, and not to the number of verb senses. Our sample verbs appear to have more SCFs on average than the whole of the French verbal lexicon.

Our relatively low recall rate is perhaps also due to our conception of the SCFs, since we did not initially wish to conflate reflexive SCFs with their non-reflexive counterparts. Given the heterogeneous nature of reflexive clitics discussed in section 2.2., we treat every frame with a reflexive clitic as an SCF distinct from the non-reflexive frame. This strategy works well for cases in which there is no equivalent non-reflexive structure, but, as one of our evaluators notes, it fails to capture the common sense of *laisser* mentioned in section 2.2. This case is an example of an SCF our system proposes as reflexive, but for which there is no obvious linguistic motivation other than frequency for distinguishing reflexive from non-reflexive frames. In such a case the reflexive frame would have been ideally subsumed under a non-reflexive frame that would take both reflexive and non-reflexive realizations into account. Thirty-one of our 176 frames include a reflexive clitic, and we conclude that the problem of how to best deal with reflexive verb entries for French could greatly affect recall rates.

4. Discussion

It is natural to ask if our results can be generalized to verbs not included in the VISL subcategorization lexicon. We find that the high precision of our results does not depend on the initial parser subcategorization lexicon. On the subset of 24 verbs not initially in the VISL subcategorization lexicon, the precision of our system is 84.2%. System precision is 87.6% on the 80 verbs included in the VISL lexicon. At these small sample sizes the difference in these results is statistically insignificant. In future experiments on verbs not in the VISL subcategorization lexicon, we expect to see equally high precision rates.

A potential concern with our results is the extent of the coverage of the frames. Although the acquisition of 27 unique frames and 176 verb-frame combinations necessarily constitutes an incomplete verbal lexicon for French, these results are nevertheless highly significant. Since SCFs, like words, are distributed in corpora according to Zipf's law (see table 1), we can estimate what percentage of the total cumulative distribution of SCFs in French our frames account for.

Using Zipf's law we predict that 85.4% of SCFs seen in corpora will be included in our results, assuming all our SCFs are grammatically plausible. Zipf's law states that the frequency of a word (or SCF, in our case) is inversely proportional to its rank. Formally, we note that

$$f \sim \frac{1}{n^{\theta}}$$

where n is the rank, f is the relative frequency, and θ characterizes the distribution. We estimate θ to have a value of 1.33, a figure arrived at by fitting the curve of the power-law distribution of our initial 27 SCFs. We do not know the total number of SCFs for French, but we can obtain an estimate using an upper bound for this figure for English. The highest number of SCFs we have seen reported for English is 160 (Briscoe and Carroll, 1997); this figure becomes our total number of unique SCFs. Figure 2 shows that more than 85% of all SCFs in French should be counted among our frames.

A methodological issue the current work addresses is the utility of binomial hypothesis testing (BHT), recently put into question in obtaining SCFs from corpus data. (Briscoe and Carroll, 1997) note that binomial filtering is the least effective step in their method of extracting SCFs from corpora. (Korhonen et al., 2000) notes that maximum-likelihood estimation (MLE) outperforms BHT. (Kilgarriff, 2005) concludes from these latter studies that BHT is inappropriate for the acquisition of SCFs from corpora. Yet (Brent, 1993), (Manning, 1993), (Sarkar and Zeman, 2000) and the current work have high precisions using BHT as a filtering method.

The difference in results in using BHT appears to lie in choosing a confidence level that is appropriate given the sample size, i.e. the number of occurrences of a particular verb. A breakdown of confidence levels, sample sizes, and resulting precision for different experiments is given in table 2. It is not our aim here to compare raw results of various works, since evaluation methods for these experiments differ as does the language worked on. Rather we wish to point out large discrepancies in precision, like those seen in table 2, with respect to confidence levels and sample sizes. From this comparison we conclude that BHT is indeed quite robust when the proper confidence level is chosen for the sample size. BHT does require the confidence level to be set empirically, yet this is the same for the MLE method in Korhonen et al. In contrast to their MLE method, Korhonen et al. appear to set the confidence level for BHT categorically at $.05$. In summary, both the BHT and MLE methods have an empirical aspect to them, and we cannot conclude from the results of (Korhonen et al., 2000) that BHT is an inappropriate method for acquiring SCFs from corpora.

5. Related Work

Extensive manual work has been done on French SCFs, covering approximately 5,000 verbal entries, ((Gross, 1975), (Boons et al., 1976), and (Guillet and Leclère, 1992)). Recently, (Gardent et al., 2005) seek to render this lexical information useful in NLP applications such as parsing. This information is more fine-grained than the present work in that it contains semantic information about the subject and complements of a given verb, such as selectional restrictions and co-indexation information for raising and control verbs. If our work errs on the side of relatively low recall at the expense of precision, (Gardent et al., 2005) note that their work risks overgenerating SCFs.

We see Gardent et al.'s approach as complementary to ours in four respects. First, the fine-grained information provided by these authors may not be optimal in certain NLP applications, at which point our coaser-grained SCFs would be advantageous; in other applications their detail might be desirable. Second, we have obtained our data from a different source than Gardent et al., which supposes our set of resulting frames will differ in some respects from theirs. In merging the resulting frames from both projects we will obtain an even more accurate lexicon with still larger coverage. Thirdly, corpus work can give helpful frequency information for a given SCF, but it cannot state which SCFs are ungrammatical for a given verb, as Gardent et al. could presumably do. Finally, the work of Gardent et al. provides a solid base for a large-coverage verbal lexicon. However, it may not be able to account for archaic or technical verbs, or particular SCFs used only in these non-standard varieties. Basing our work on corpus data, we can account for this "periphery" of the French verbal lexicon. For example, with the date-constraining feature of Frantext, we could limit our corpus to texts occurring before 1700 and thus extract archaic SCFs. If given corpora from technical fields, we could also extract SCFs that are unique to a certain domain as well as SCFs for rare and technical verbs.

6. Conclusions and Future Work

This paper presents a method of automatically extracting French subcategorization frames (SCFs) from corpora using binomial hypothesis testing. We obtain high levels of type precision (86.8%) at a decent token recall rate (54.3%). We conclude that our results in using binomial hypothesis testing are robust because our confidence level was appropriate for our sample size. The work of (Korhonen et al., 2000) does not yield conclusive evidence against binomial hypothesis testing, since in their study confidence levels were not appropriate for the sample size. However, as Korhonen et al. demonstrate, other methods can also be effective in determining SCFs.

We would like to investigate the utility of SCFs for lexicalized probabilistic parsers in greater detail. Given that syntactic arguments have freer word order in French than in English, would subcategorization information for French be helpful beyond the distance information used by current state-of-the-art lexicalized parsers? Is lack of subcategorization information one reason for which parser accuracy for French is significantly lower than for English? Answers to these questions will help to determine the optimal parsing configuration (e.g., distance vs. subcategorization) for a given language.

7. References

A. Abeillé, L. Clément, and F. Toussenel, 2003. *Building a treebank for French*, pages 165–188. Treebanks: Building and Using Parsed Corpora. Kluwer Academic Publishers.

A. Arun. 2004. Statistical Parsing of the French Treebank. Master's thesis, University of Edinburgh.

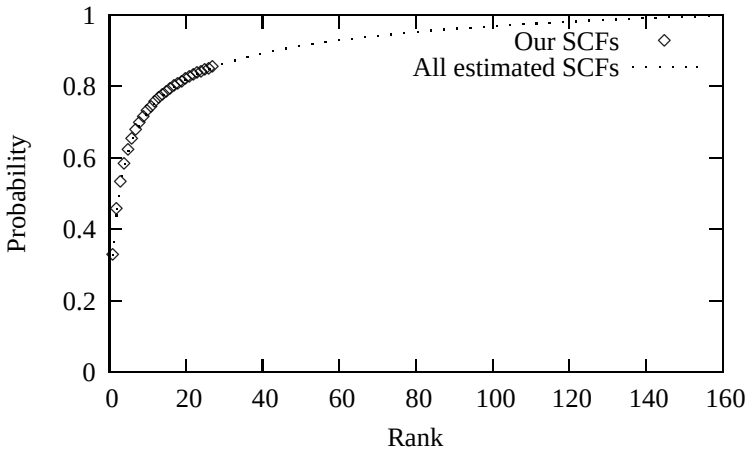

Figure 2: An estimate of the cumulative density function of SCFs in French. Note that the current results should account for 85.4% of SCFs seen in corpora.

Work	Sample size	Confidence level	Precision
(Brent, 1993)	50-150	.02	96-100%
(Manning, 1993)	?	.02	90%
(Briscoe and Carroll, 1997)	$< 1,000$	.05	65.7-76.6%
(Sarkar and Zeman, 2000)	?	.05	82-88%
(Korhonen et al., 2000)	average of $3,000$	.05	50.3%
Current work	40-125	.02	86.8%

Table 2: Various studies' precision results as they relate to sample sizes and confidence levels. Here sample size is the number of occurrences of one verb, and a "?" indicates that this number was not indicated in the paper. Recall rates are not mentioned as some works do not discuss the method for determining recall, e.g., token vs. type.

E. Bick. 2003. A CG & PSG Hybrid Approach to Automatic Corpus Annotation. pages 1–12. Proceedings of Shallow Processing of Large Corpora (SProLaC 2003).

J.-P. Boons, A. Guillet, and C. Leclère. 1976. *La structure des phrases simples en français: Constructions intransitives*. Droz, Geneva.

M. Brent. 1993. From grammar to lexicon: Unsupervised learning of lexical syntax. *Computational Linguistics*, 19(2):243–262.

E. Briscoe and J. Carroll. 1993. Generalised probabilistic LR parsing for unification-based grammars. *Computational Linguistics*, 19(1):25–60.

E. Briscoe and J. Carroll. 1997. Automatic Extraction of Subcategorization from Corpora. pages 356–363. Proceedings of the 5th ACL Conference on Applied Natural Language Processing (ANLP).

P. Chesley and S. Salmon-Alt. 2005. Le filtrage probabiliste dans l'extraction automatique des cadres de sous-catégorisation. ATALA Workshop, 12 March, Paris.

M. J. Collins. 1999. *Head-driven statistical models for natural language parsing*. Ph.D. thesis, University of Pennsylvania.

D. Estival and S. Lehmann. 1997. TSNLP — Des jeux de phrases-test pour l'évaluation d'applications dans le domaine du TALN. *Traitement automatique des langues*, 38(1):155–171.

C. Gardent, B. Guillaume, G. Perrier, and I. Falk. 2005. Extracting subcategorisation information from maurice gross' grammar lexicon. *Archives of Control Sciences*, 15(3):253–264.

M. Gross. 1975. *Méthodes en syntaxe: régime des constructions complétives*. Hermann, Paris.

A. Guillet and C. Leclère. 1992. *La structure des phrases simples en français: Constructions transitives locatives*. Droz, Geneva.

A. Kilgarriff. 2005. Language is never, ever, ever random. *Corpus Linguistics and Linguistic Theory*, 1(2):263–276.

A. Korhonen, G. Gorrell, and D. McCarthy. 2000. Statistical filtering and subcategorization frame acquisition. pages 199–205. Proceedings of the Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora.

C. Manning. 1993. Automatic Acquisition of a Large Subcategorization Dictionary from Corpora. pages 235–242. Proceedings of the 31st Association for Computational Linguistics (ACL 1993).

L. Romary, G. Francopoulo, and S. Salmon-Alt. 2004. Standards going concrete: from LMF to Morphalou. Computational Linguistics (COLING) workshop.

A. Sarkar and D. Zeman. 2000. Automatic extraction of subcategorization frames for Czech. pages 691–697. Proceedings of the 18th conference on Computational Linguistics (COLING).