

A Note on a Study of Cases

K. Sparck Jones and B. Boguraev
Computer Laboratory, University of Cambridge
Corn Exchange Street, Cambridge CB2 3QG, England
sparckjones@cl.cam.ac.uk

December 1985

Memo, KSJ August 1997

This document is an enlarged version of the Note published in 'Computational Linguistics' 13, 1987, 65-68; it formed the first part of 'Material concerning a study of cases', Technical Report 118, Computer Laboratory, University of Cambridge, 1987; the second part of the Report was made up of two detailed data files named woodsent/full1, where example sentences are ordered by preposition, and woodsent/sort1, where they are ordered by case label.

Note that the source sentences in the two data files are taken from F.T. Wood, 'English prepositional idioms', Macmillan, London, 1967 (ELBS printing, 1975) and are reproduced with the permission of the author and publisher. The material in the files is made available only for bona fide research purposes, and in any publication referring to it explicit acknowledgements should be made citing author, title and publisher for the source material, and ourselves for the additional material and distribution.

The present version of this document has been reformatted from the antique original.

Like many others, we have used a language analyser which builds meaning representations expressing semantic case roles; specifically, Boguraev's (1979) analyser builds dependency trees with word senses defined by *semantic category primitive* formulae, and with case labels, *semantic relation primitives*, on the constituents of verb (and some other) structures. The category primitives, which are taken over from Wilks, are part of the analysis apparatus, as well as of its output. The relation primitives figure primarily in the output of analysis, derived from many sources, including conventional syntax; they have no systematic relation with Wilks' 'case-type' category primitives, and achieve a much more thoroughgoing semantic labelling of sentence structure than anything envisaged by Wilks (Wilks 1977).

Using the analyser for more varied and demanding purposes than Boguraev's original tests (see e.g. Boguraev and Sparck Jones 1983) left us dissatisfied with the original set of case relations. We therefore carried out a detailed analysis of a large sample of English sentences

to evaluate our proposals for a better-founded and more comprehensive set of case relations. This study exploited F.T. Wood's 'English prepositional idioms' (Wood 1967), which provides a careful account, supported by extensive examples, of the uses of English prepositions and preposition-like terms. For instance:

WITHIN

(1) Inside

Within the house all was quiet.
The Kingdom of God is within you.

(2) Amongst the members of a group.

Opinion within the profession is divided.

(3) Inside specified bounds or limits.

They were ordered to remain within the precincts of the college.
The scholarship is open to anyone residing within fifty miles
of the university.
He always strove to live within his income.

.....

Our study was intended to establish both the justification for each case relation individually, by reference to a range of sentences, and the plausibility of the set of relations as a whole, by reference to the complete set of sentences. Looking at Wood's description of a preposition's sense, and its accompanying illustration(s), we tried to assign a case label to the link between the sentence elements made by the preposition which we felt captured the essential nature of that link, at the level of generality represented by a set of 20-30 cases. Thus 'location' would be the label associated with a number of specific space-relation prepositions, eg "above", "at", "by". The study was primarily concerned with prepositionally-based cases, but we considered other case instantiations, though not systematically. While additional cases might therefore be motivated by non-propositional sources we have not examined, we believe that a complete set would have to include something like our list, to cover prepositions adequately.

The study was a purely investigative one: we were interested in the legitimacy of the cases as characterisations of sentence relations, and did not address the question of how the specific assignments, for the individual sentences, could be achieved automatically. However as we had already demonstrated that a quite refined set of cases could be applied by Boguraev's existing analyser, and have since, as practical need has arisen, implemented further cases, we feel some confidence in the feasibility of automatic assignment of the cases in the present set. (This of course accepts that some individual sentences may present considerable difficulties.)

In the exposition which follows we therefore include brief notes, as appropriate, of the kinds of syntactic and semantic sources we would expect an analyser combining conventional syntax with some Wilksian semantics to handle.

We are offering our results here not with any great claims about the especial novelty or merits of our case set, but rather as a practical contribution to the utilisation of case labels, justified by the large and varied sentence sample studied. Though domain-specific strategies may reduce the need to handle lexical ambiguity, especially in nouns and verbs but even in prepositions, it may nevertheless be necessary even in the domain-specific case to provide for distinct semantic functions in prepositions.

The remainder of this note therefore consists of

- (1) the list of our cases, with descriptive notes and illustrations;
- (2) extracts from the complete sentence set, showing
 - a) the treatment of all the senses of some prepositions, and
 - b) all the applications to different prepositions of some cases.

Complete listings for the full sentence set are available from the authors. We acknowledge the immense value of Wood as a source. We know there are some uncorrected errors in the sentence listings. This material was prepared in 1981. Since then the analyser has been extended, for example to deal with determiners, including quantity terms, particularly as these might be expressed as quantifiers. The treatment of quantities here is confined to those expressed through prepositions; however our experience with quantity terms has led us to make a small number of changes to the original 'quantity' case examples in our listings. Our ideas on states have also changed, and our treatment of 'state' in the sentence listings is not necessarily what it would be now. However as the analyser does not tackle the 'state' case in other than very limited forms, and is certainly not comprehensive enough to deal with most of the examples in the sentence listings, we have left these unchanged.

We thus offer the data essentially in the modest hope that it may, in spite of its imperfections, be of some general practical utility, not least in drawing attention to the complexity of unrestricted natural English sentences and to the problems of providing convincing case-based representations for them.

(1) THE CASE LIST

In the examples each case, *ortag*, name is prefaced by its abbreviation. The meaning of the case is not explicitly defined, but is assumed to be ostended by the subsequent illustrations (and more particularly by the full sets of example sentences derived from Wood). Sources of the case are then indicated, for example the lexical entries for prepositions, or those for verb and adjective senses with characteristic (even idiomatic) accompanying prepositions, all of which embody contextual patterns; or the sources may be syntactic structures like embedded clauses and complements (our notes do not necessarily exhaust the possibilities here): these items are prefaced by "program", as these structures are identified by the analyser. Syntactic entities, ie the contents of particular ATN registers exploited by an analyser like ours, are not

always correlated with particular cases: the correlation is context-dependent. Similarly, Wilks 'case-type' subformula primitives like OBJE may suggest, but are not necessarily correlated with, particular cases, and so are only indicated for situations where the correlation is very likely. The normal linkage marked by a case is between verb and noun group, but two nominal groups may also be case-linked, and also more than two items (though this is not common). In the illustrations we have adopted the convention of marking the head words of the two linked constituents by -. Thus the tag 'after' labels the relation between "left" and "breakfast" in "John left following breakfast", written as "John -left following -breakfast". In the cases where complete constituents like embedded clauses fill case roles, they are bracketted with [], and the whole item is marked with -. Those cases or sources we have not yet programmed are marked with a +.

There are 28 tags altogether. They are alphabetically ordered here on their abbreviations.

(acc) ACCOMPANIMENT

dictionary: eg "with"
can link nominals

John --went to the zoo (along) with --Mary.
John --went everywhere with his --violin.

(act) ACTIVITY

dictionary: eg "at"
program: embedded clauses

John --beat Mary at --chess.
Peter was --top of the class in --history.
Aloysius --beat Sebastian --running.
John --beat Peter in all their --examinations.
John --escaped (from) the --conflict.

(adest) +ABSTRACT-DESTINATION

dictionary: eg "to"

I --reached my --conclusion.
When heated, water --turns into --steam.

(aft) AFTER

dictionary: eg "after"
program: ing-phrases, +adverbials
can link nominals

John --left after --noon.
John --left after --Bill.

Peter --left following --breakfast.
Mary will come the --day after --tomorrow.
John, --[having bought the book], --took it home.

(ag) AGENT

program: contents of "subj" register
Wilks' primitive SUBJ may lead to agent

--He --bound the books.
--Mary --walked to town.
--Books --stimulate me.
John was --hit by --Bill.
--Malaria --killed the girl.
--Unemployment --depresses everyone.

(aloc) +ABSTRACT-LOCATION

dictionary: eg "in"

Parry was --fixed in his --mind that the Mafia was out to get
him sometime.
The four friends --shared the work between --them.

(asour) +ABSTRACT-SOURCE

dictionary: eg "from"

I --got them from --Bill.
The conclusion --followed from the --premises.
She --makes dresses of --silk.

(attr) ATTRIBUTE

dictionary: eg "with"
only links nominals (but perhaps in unobvious dependency)

The --girl in --blue was happy.
The --girl with the pink --hat was sad.

(bef) BEFORE

dictionary: eg "before"
+program: adverbials
can link nominals

John --left before --noon.

The stock market --was very active ahead of the --Budget.
Aloysius got married the --day before --yesterday.

(comp) +COMPARISON

dictionary: eg "as"
program: ?
can link nominals

John --passed as --Bill.
Mary --feels like a new --woman.
He was given --cash instead of --kind.
Time --flies like an --arrow.

(dest) DESTINATION

dictionary: eg "to"

John --went to --Paris.
Bill --walked up to --John.
John --marched into --town.

(dire) DIRECTION

dictionary: eg "down"
John --walked after --Bill.
John --ran down the --hill.

(force) FORCE

dictionary: eg "of"

The girl --died of --malaria.
The girl --died from an --accident.

(goal) GOAL

dictionary: eg "for"
program: to-complements

He --walked for --pleasure.
He --worked hard for his --money.
John --shot at --Mary.
John --went to town in order --[to buy a shirt].

(inst) INSTRUMENT

dictionary: eg "by"
+program: complements

The boy --caught the fish with a --minnow.
John --went to Paris by --plane.
John --sought visions through --drugs.
John --sought visions with the --help of drugs.

(loc) LOCATION

dictionary: eg "at"

Bill --lived in --Paris.
John --walked in the --park.
The --girl at --Smith's was not helpful about buying a book
for a dyslexic child.
Sebastian --felt pain in his --foot.
John --felt pain behind his --ear.

(man) MANNER

dictionary: eg "with"
program: embedded clauses, +adverbs
They --ran with --speed.
John --did it --gladly.
Peter --won the race by --cunning.
John --argued in his own --way that he was being persecuted.
He --did it by --[working very hard].

(mobj) MENTAL-OBJECT

dictionary: eg "about"
program: embedded clauses, to-complements, that-complements

John --thought of --fish.
John --pretended --[he was a dragon].
John --planned --[to go to London].
John --threw light on the --problem.
John --told Mary --[that he believed her].
John --thought about --fish.
They --talked about --politics.
John --asked about the --fish.

(obj) OBJECT

(almost a default)
program: typically contents of "obj" register

(not necessarily correlated with OBJE)

John --bought --books.
John --threw --light on the problem.
--Books were --bought by John.

(poss) POSSESSED-BY

program: noun phrases (+limited treatment)
only links nominals

The --daughter of the --Mayor was blonde.
The --pocket of my uncle's --coat was empty.
The --leader of the --men was bald.

(quant) QUANTITY

program: noun phrases
can link nominals

There were four --students including a --dwarf.
It was too --long by twelve --inches.

(reas) REASON

dictionary: eg "of" (re predicate adjectives)
+eg "because of" (re nominals)
program: embedded clauses

John is --afraid of --spiders.
John is --afraid of --[being apprehended by the police].
John is --happy about --school.
John --drove slowly because of the --fog.

(rec) RECIPIENT

dictionary: eg "to"
program: contents of "obj" register

They --bought tickets for --her.
Susan --gave the flowers to --Jill.
John --kissed --Mary.
John --killed --Mary.
John --shot --Mary.
Mary made large --donations to --charity.

(sour) SOURCE

dictionary: eg "from"

She --took some money out of her --bag.
I --got them from --Paris.

(state) STATE

+dictionary eg "of"

program: predicate adjectives

Wilks' primitive KIND tends to lead to state

His conduct --is --admirable.
The apple --appeared --red.
He wears --green --trousers.
The apple --turned --red.--

(subj) SUBJECT

(almost a default)

program: various places

(not necessarily correlated with Wilks' primitive SUBJ)

--Books --are nice.
John is --fascinated by --books, especially ones published by
private presses.

(tloc) TIME-LOCATION

dictionary: eg "at"

They --ate breakfast at --noon.
John --runs in the --morning.
John --runs on --weekdays.
I --wrote my thesis during --1979.
John --sang during his --run in the park.

(tspan) TIME-SPAN

future dictionary: eg "throughout"

program: embedded clauses, ing-phrases

It --took me six --months to do it.
I --was a bus conductor for two --days.
Throughout --June it --rained every day.
John --sang --[running in the park].
John --slept during the --concert.

--[Running], Aloysius --beat Sebastian.

(2) SENTENCE EXAMPLES

These sentences, taken from Wood, cover all the straightforward senses of the prepositions, as numbered by Wood, with one sentence per sense. Each sentence is thus representative of the set of examples given for each sense by Wood; it is usually a direct copy of one of Wood's, though occasionally sentences have been simplified (primarily to make them shorter) or have been supplied to fill out Wood's remarks. Items explicitly described as "Phrases" in Wood, or obviously idiomatic expressions, have not been included.

In the full set of example sentences * is used to flag items described by Wood as not strictly prepositions, but functioning as such, eg "concerning"; () flags items described by Wood as obsolete, eg "betwixt", and ? flags items not deemed by us to be regular prepositions because they are idioms, postverb prepositions, or non-prepositions, eg "on" in "on the make".

The constituents linked by the case relations are marked by –, and possibly also by [], as in the previous section; each sentence is preceded by the relevant case label, in its abbreviated form. Some sentences have more than one tag, for alternative readings.

In the complete sentence sample there are 421 sentences. The average number of sentences per case is 15, with low variation. However 'location' has 80 sentences, reflecting the very large number of different space-relation prepositions there are. It is possible that more specific space-location cases would be desirable, though the right level of discrimination and particular set would not be easy to establish. We are in any case not suggesting that the use of a case tag in the representation of a sentence delivered by the analyser makes it unnecessary to indicate the specific lexical sense of the preposition.

(a) Extract from the complete sentence sample showing the full treatment of representative prepositions

ABOUT

/mobj/ I --know nothing about the --matter.
/subj/ There is --something about --him that I do not like.
/loc/ She --had a gold chain about her --neck.
/acc,loc/ She is a woman who likes to --have children about --her.
/loc/ Do you happen to --have his letter about --you?
/loc/ The papers were --scattered about the --floor.
/loc/ He --goes about the --country lecturing on the evils of drink.
/loc/ Mrs Jones does not appear to --be about the --house.
/??/ I was about to speak when my companion distracted me.
/act/ I do not like being interrupted when I --am about an
important --task.

AMONG, AMONGST

/loc/ They live in that --house among the --trees.
/loc/ I --found the letter amongst some --papers
/dest,rec/ The food was --distributed amongst the --poor.
/asour/ They couldn't --raise five pounds among --them.

AT

/loc/ Mr Bwown --is at the --office.
/tloc/ The concert --starts at half past --seven.
/loc/ At the second --roundabout, --turn left.
/dest/ We have --arrived at our --destination.
/act/ The two tribes --were constantly at --war.
/act/ She loved to watch the --children at --play.
/man/ He --left at a moment's --notice.
/man/ He --vaulted over the gate at a single --leap.
/force/ I have --called to see you at the --request of a friend.
/man/ Cars are --parked at their owners' --risk.
/man/ The car --whizzed along at sixty --miles an hour.
/force/ She felt sure she would --faint at the --sight of blood.
/dest/ As we passed the gate the dog --flew out at --us.
/act/ That boy --is clever at --mathematics.

BY

/ag/ The new library was --opened by the --Mayor.
/inst/ I shall --go by the 10.30 --train.
/obj/ He --seized me by the --arm.
/man/ He --succeeded by hard --work.
/man/ I --know him by --sight.
/loc/ She --sat by the --pillar.
/loc/ We --went by the --shop without realising it.
/bef/ They should have --been here by --now.
/tspan/ They decided to --travel by --night.
/quant/ We --lost the match by one --goal.
/reas/ She --was very generous by --nature.
/inst/ They always --addressed each other by their Christian --names.
/loc/ I --went from London to Machester by --Sheffield.
/inst/ He --swore by all the --Gods he would conquer.
/attr/ He brought a --person by the --name of Smoth.
/quant/ The carpet is three --yards by --four.
/??/ The box is too heavy for me to lift by myself.
/quant/ --Little by --little his savings accumulated.
/quant/ Milk is --sold by the --pint.
/inst/ By my --watch, it'--s time for lunch.
/force/ The meeting will be --held in the school, by --permission o
the headmaster.

(b) Extract from the complete sentence sample showing the full set of sentences for representative cases

/acc/ I --believe with --Whitman that self-expression is all.
/acc/ I am sending you his --letter, together with my --reply.

/acc/ I will --dance with --you.
/acc/ Plant --disease, together with a prolonged --drought, has ruined
the crops.
/acc/ The children --came without their --parents.
/acc/ There are other --people to be considered besides --you.
/acc/ Would you like to --come to the theatre with --us.
/acc,loc/ She is a woman who likes to --have children about --her.

/aloc/ --Opinion within the --profession is divided.
/aloc/ A duke --comes before a --viscount.
/aloc/ All our difficulties --are now behind --us.
/aloc/ As a scholar, he --is far above --me.
/aloc/ Don't let such a trivial matter as this --come between --us.
/aloc/ He --has nearly fifty men under --him.
/aloc/ He --has no malice in --him.
/aloc/ He --is on the --Board.
/aloc/ House prices are not likely to --fall below the present --level.
/aloc/ I am not in charge: I --have two people over --me.
/aloc/ Inside the --party there --is disagreement on policy.
/aloc/ My eldest son --is in the --navy.
/aloc/ She --is rather behind the --rest of the class.
/aloc/ The offender was --brought before the --magistrates.
/aloc/ The other Elizabethan dramatists --are far beneath --Shakespeare.
/aloc/ There had --been a lifelong friendship between --them.
/aloc/ There may --be nothing in the --rumour.
/aloc/ They --shared the work between --them.
/aloc/ We'--ve a difficult task in front of --us.
/aloc/ You will --find them under --'Biography'.

/inst/ By my --watch, it'--s time for lunch.
/inst/ He --looked at it through his --binoculars.
/inst/ He --swore by all the --Gods he would conquer.
/inst/ I managed to --beat the dog off with a --stick.
/inst/ I shall --go by the 10.30 --train.
/inst/ Many locomotives nowadays --run on --oil.
/inst/ Marian Evans --wrote under the --name George Eliot.
/inst/ The parcel was --wrapped in --paper.
/inst/ The poem has been --set to --music.
/inst/ The prosecution was --brought under the --Vagrancy Act.
/inst/ The repairs --cost me over five --pounds.
/inst/ The roof of the building is --supported on --pillars.
/inst/ They always --addressed each other by their Christian --names.
/inst/ You can --use that saucer for an --ashtray.

/man/ Cars are --parked at their owners' --risk.
/man/ He --is, without --doubt, a very able person.
/man/ He --left at a moment's --notice.

/man/ He --said nothing in --reply to my questions.
/man/ He --succeeded by hard --work.
/man/ He --vaulted over the gate at a single --leap.
/man/ I --did it without --thinking.
/man/ I --know him by --sight.
/man/ I only --said it in --fun.
/man/ It --is all right in --theory.
/man/ She --did the work to the --best of her ability.
/man/ She --lay back with her --eyes closed.
/man/ The car --whizzed along at sixty --miles an hour.
/man/ The child --screamed in --terror.
/man/ The last few years of his life were --spent in --poverty.
/man/ The water --shot up in a --fountain.
/man/ The work has been --carried out according to your --instructions.
/man/ They --played out of --tune.
/man/ They let him --have it on --loan.
/man/ To our --surprise, the train --was early.

References

B.K. Boguraev, *Automatic resolution of linguistic ambiguities*, Technical Report 11, Computer Laboratory, University of Cambridge, 1979.

B.K. Boguraev and K. Sparck Jones, 'How to drive a database front end using general semantic information', *Conference on Applied Natural Language Processing*, Santa Monica, 1983, 81-88.

Y. Wilks, 'Good and bad arguments about semantic primitives', *Communication and Cognition*, 10, 1977, 181-221.

F.T. Wood, *English prepositional idioms*, London: Macmillan, 1979.