

# Automatic extraction of lexical-semantic relations from dictionary definitions

Thierry Fontenelle

## Introduction

The term *reusability of lexical resources* has now become a buzzword and much attention is being paid to how lexical resources (and more particularly computerized and machine-readable dictionaries) might be reused and incorporated into real-life natural language processing systems. It has even become one of the current research areas in computational linguistics and the European Community has recently decided to support and launch a research programme aiming at investigating which existing dictionaries and/or terminological databanks would be reasonable candidates for importation into machine translation, question-answering systems or natural-language front-ends.

Since the construction of dictionaries for MT and NLP is a highly time-consuming task and since parsing systems require substantial lexicons, we start from the assumption that it would not be desirable to code lexical items manually. Various studies (in Liège and other places) have already shown that some existing computerized dictionaries contain a lot of the information that is essential for those NLP systems. The *Longman Dictionary of Contemporary English* (henceforth, *LDOCE*, 1978) for example can be used as a starting point to extract a wide variety of lexico-semantic relations, exploiting the various kinds of coded information that its file provides:

- semantic codes (i.e. cooccurrence restrictions with a set of semantic markers à la Katz and Fodor);
- subject field codes (pragmatic information);
- syntactic information, in the form of grammatical codes which describe the environment in which a lexical item can be inserted.

All these codes refer to formalized information and can therefore be relatively easily retrieved once the dictionary has been organized in proper database format. Moreover, a dictionary is not free from imperfections and inconsistencies. We believe that it is possible to correct and improve the lexical database by looking at some pieces of information which have not yet been exploited on a large scale because they are not sufficiently formalized to be used in computer systems. Definitions, for instance, can be reused and tapped since they encapsulate much of the information that is needed for word sense assignment and disambiguation.

In this paper, we concentrate on a computationally relevant subset of lexical items extracted from the *LDOCE* database, viz ergative verbs, i.e. verbs which are both transitive and intransitive and where the subject of the intransitive construction may be the object of the transitive construction, as in «*the enemy sank the boat*» vs. «*the boat sank*».

### Ergativity

We have already argued that ergativity is translationally relevant since English ergative verbs may be translated into French in three different ways (for more details, see Fontenelle and Vanandroye, 1989):

– no change at all:	<i>John cooks a rabbit</i>	vs.	<i>a rabbit cooks</i>
	<i>Jean cuit un lapin</i>	vs.	<i>un lapin cuit</i>
– reflexive verb:	<i>John opens the door</i>	vs.	<i>the door opens</i>
	<i>Jean ouvre la porte</i>	vs.	<i>la porte s'ouvre</i>
– causative "FAIRE":	<i>John boils water</i>	vs.	<i>water boils</i>
	<i>Jean fait bouillir de l'eau</i>	vs.	<i>l'eau bout</i>

In the same study we showed that ergativity is lexically governed and that this property should be coded at definition level if we want our computerized lexicon to account for the transitive and intransitive usages of this set of verbs. We also showed that the definition patterns (the so-called defining formulas, to quote Ahlsvede and Evens, 1987) provided evidence that a verb was ergative: using the predicate decomposition approach developed by the generative semanticists, we were able to decompose the meaning of ergative verbs into a set of semantic primitives (atomic predicates) where the optionality of the primitive CAUSE accounts for the transitive/intransitive alternation (also called causative/inchoative alternation). Indeed, a verb such as «open» can be decomposed into two sets of predicates:

- (1) open = CAUSE to BECOME OPEN (e.g. John opens the door);
- (2) open = BECOME OPEN (e.g. the door opens).

We see that the former construction (the transitive one) makes use of the primitive CAUSE (to be distinguished from the English verb «cause»). We also notice that an ergative verb actually opens slots for two arguments. Typically, the general structure of such a verb is as follows:

(X) \_ Verb \_ Y

where X is the AGENT and Y is the PATIENT. The term AGENT is used here in a broad sense and refers to the cause of the action: it may represent an animate being, an initiator (as in «the officer marched the soldiers») or a force (as in «the winter wind froze the pond»). Some justification for conflating these concepts under the more general heading of AGENT can be found in Allerton, 1982.

In the structure above, the agent is optional and the patient is the subject when there is no agent (intransitive construction). When the agent is mentioned, the patient is the object and the verb is transitive. It should be noted that this alternation is typical of change-of-state verbs (for further details, see also Levin, 1988).

In our preliminary research, we had decided to extract what we thought to be an exhaustive list of ergative verbs, starting from the idea that those verbs, although not coded explicitly, would be located using a conjunction of formalized or semi-formalized criteria, namely a combination of particular grammatical codes

and of definition patterns. We therefore concentrated on the various syntactic codes which express transitivity and intransitivity: the list of codes which had to be assigned at definition level is to be found in Fontenelle and Vanandroye 1989. As for the definition patterns, we listed the possible surface realizations of the semantic primitive CAUSE. Those realizations were to be found within the definitions of a verb for it to be ergative:

- to (cause to): as in the entry «tighten (1)»;
- to (allow to): as in the entry «ooze (1)»;
- to (help to): as in the entry «pull through (2)»;
- to make or become: as in the entry «shorten (1)»;
- to bring or come: as in the entry «land (1)».

It will be noticed that the causative/inchoative alternation is described by the lexicographer's use of parentheses in the first three cases whereas the disjunctive conjunction «or» takes care of this alternation in the last two definition patterns. Another point is that «make» can actually be decomposed into «CAUSE to BECOME» and «bring» into «CAUSE to COME».

The fact that the *LDOCE* lexicographers have used a restricted vocabulary of around 2,000 items to write their definitions is of great help since this practice has imposed some constraint on the defining style: variations being somewhat limited, it is easier to locate ergative verbs once we have detected valid predictors of ergativity.

### Dictionaries and consistency

A huge problem remains when one tries to exploit the regularities found in a dictionary: hand-crafted lexicons are not always reliable insofar as they are quite often marred by inconsistencies, errors of omission and errors of commission (on this crucial issue, see Boguraev and Briscoe, 1989; and Michiels, 1982). Since no checking procedure was attempted, we must not expect any algorithm for retrieving information to be one hundred per cent reliable. Therefore, our aim was to find other criteria which would provide us with clues as to the ergative property of a verb. Expanding on some recent work on the relational models of the lexicon (see i.a. Calzolari, 1988b), we hypothesized that there is a relationship that links derivational morphology with the lexical-semantic representation of lexical items. The analysis of four suffixes enabled us to locate and cope with various kinds of errors and inconsistencies in the dictionary. The suffixes we examined are the following:

- EN: as in the entry «quieten (1)»;
- ATE: as in the entry «precipitate (3)»;
- FY: as in the entry «liquefy (1)»;
- IZE: as in the entry «demobilize (2)».

Those four suffixes are, at least to some degree, predictors of ergativity: they can be attached to some base form to derive a verb which sometimes displays the transitive/intransitive alternation. It is of course not the case that all verbs ending in -EN, -ATE, etc. are ergative, but a manual analysis of such verbs enabled us to locate

nearly 20 ergative verbs which were not retrievable otherwise (e.g. «*quicken*» which is only coded [T1] for «transitive» and where the example displays typical intransitive usage).

### Lexical regularities

As Levin (1988) points out, the causative/inchoative alternation is typical of change-of-state verbs (see also Atkins *et al.*'s discussion of «bake», 1988). Such a generalization is very important since it could dramatically reduce the amount of information needed at entry or definition level. Knowing in advance that a verb expresses a change of state would spare us the trouble of explicitly specifying the subcategorization frame of that verb. We would thus avoid quite a large number of redundancies and hence reduce the size of lexical entries. From a more practical point of view, we are interested in retrieving a list of ergative verbs. The first list we produced (using the combination of grammatical codes and definition patterns mentioned above) was examined carefully (ca. 750 items). We were able to detect some regularities and we noticed that a fairly sizeable number of those verbs belonged to two semantic fields, namely verbs expressing a change of colour and verbs expressing the production of a sound. These are actually sub-classes of change-of-state verbs and we want to show that it is possible to identify such verbs if we discover a set of key-words. This way of tackling the problem proves to be of considerable value in at least two respects:

- a. It enables us to detect inconsistencies in the treatment of these verbs and hence to improve and refine our list of ergative verbs (finding new items which had not been found in our previous study).
- b. More importantly, it sheds light on the structure of the lexicon and shows that it must not be exclusively viewed as a repository of idiosyncratic information, but rather as a highly-structured network whose nodes are closely interrelated. This approach is useful if we try to exploit the regularities that underlie the lexicon, with a view to establishing thesauric links between lexical items (see Michiels and Noël, 1982, or Calzolari, 1988a,b).

### Change of colour

Verbs expressing a change in colour typically involve an AGENT (a human being or a physical force which can act upon things) and a PATIENT (usually a concrete object which can undergo a change of colour). Most of these verbs satisfied the criteria which we had defined, namely a combination of two grammatical codes expressing transitivity and intransitivity together with a particular definition pattern involving an optional realization of CAUSE. Here is the list of these verbs, together with the name of the colour or key-word(s) with which they appear:

blacken (1)	(black)
bleach (1)	(white)
brighten (1)	(bright)
brown (1)	(brown)

char (1)	(black)
clear (1)	(clear)
cloud (3)	(dark)
crimson (1)	(crimson)
darken (1)	(dark)
dim (1)	(dim)
discolour (1)	(colour)
fade (1)	(colour)
grey (1)	(grey)
kindle (1)	(red+colour)
light (3)	(bright)
light up (1)	(bright+colour)
lighten (1)	(bright+dark)
pale (1)	(pale)
redden (1)	(red)
silver (2)	(silver+colour)
tan (2)	(brown)
tarnish (1)	(bright+colour)
whiten (1)	(white)
wither (1)	(colour)
yellow (1)	(yellow)

One first thing to note is that 6 of these verbs end in -EN (this is an application of the derivation rule where an adjective X denoting a colour is combined with the suffix -EN to produce a verb which denotes the change of state in question):

X + -EN → make or become X (e.g. *blacken*).

Another thing to note is that some verbs are not specific to a particular colour, but refer to a change in the luminosity of the patient (e.g. *pale*, *light up*, *wither*, *darken*...); hence the use of a key-word which is not precisely a colour term («pale», «bright», «dark» or simply the word «colour» itself). It should be stressed that the relationship that links a verb such as «whiten» to the adjective «white» (i.e. the link between the first and the second columns in the figure above) is just the kind of semantic relationship which we would like to account for in the thesauric organization of the lexicon that we have in mind. The computer here is considered as a tool to investigate the structure of the lexicon in which all items are linked to one another by means of pointers: the latter enable the user to generate sub-sets of words, detect the link between two such separate items as «bleach» and «white» or examine the words which belong to a given lexical field.

We now return to our primary aim, which was to generate a comprehensive list of ergative verbs. Once we have discovered that a reasonable number of verbs expressing a change of colour display the causative/inchoative alternation, we can start from the concept of change-of-colour to try to spot other verbs which pattern in a similar fashion. This approach is onomasiological since we start from the meaning to find out how this meaning can be expressed. It is fairly easy if we query all items that contain particular key-words within their definition field. This type of query must of course be performed in a semi-automatic way since we run the risk of retrieving more

items than just those we are interested in. Moreover, we have to specify that we want to locate words whose part of speech is a verb, but we do not see this as a problem once the dictionary is organized in proper database format, as is our *LDOCE* database.

We examined the following key-words:

RED - WHITE - BLACK - BLUE - YELLOW -  
GREEN - DARK - BRIGHT - COLOUR.

This approach enabled us to identify six additional verbs which had not been found previously. Let us turn to the reasons why those verbs had not been detected. The first verb we examine is COLOUR:

COLOUR (1) [T1.X1.X7]: to cause (something) to have colour or a different colour, esp. with a CRAYON or a pencil rather than a brush; (2) [I0]: to take on or change colour

We see that the *LDOCE* lexicographers have adopted the splitting strategy instead of favouring the lumping strategy which is used with all the other ergative verbs. The absence of parentheses around «cause» forced us to consider it as a purely causative verb, which it definitely is not. In our definition-centred organization of the lexicon, all our queries must be carried out at definition level. The verb «FLUSH» (definitions 4 and 5) displays a similar problem (4: to turn red; 5: to cause to become red). Other verbs, however, include information that the verb is ergative but use another type of defining formula or resort to another ergative verb as genus word. Consider the following words:

DYE (1) [T1.I0.X7.L7]: to give or take (a stated) colour by means of DYE;  
SCORCH (1) [T1.I0]: (a) to burn (part of) a surface so as to change its colour, taste or feeling... (b) (of such a surface) to burn in this way;  
SET (10) [T1.I0]: (a) to fix (a colour) against being changed as by water (b) (of a colour or colouring matter) to become fixed;  
SPOT (2) [T1.I0]: to mark or be marked with coloured or dirty spots.

DYE clearly illustrates the use of another defining formula where «give» can be decomposed into «CAUSE to HAVE». The second example shows that lexicographers sometimes use an ergative verb as genus word: to define SCORCH, they make use of the verb «burn» used transitively under sub-letter (a) and intransitively under sub-letter (b). This of course forces us to retrieve a preliminary list of verbs before tackling the problem: this sort of Catch-22 situation raises the crucial problem of bootstrapping in automatic information acquisition from computerized lexical databases (see a.o. Wilks *et al.* 1989).

The last example illustrates the use of a passive construction in co-occurrence with an active construction linked by «or» (mark or be marked). The structure of this defining pattern reads as follows: V or be V-en. Exploiting this type of practice would probably reveal other inconsistencies in the coding of ergative verbs.

### Is generalization possible?

We have already mentioned the problem of generalization which may prove crucial when a large-size lexicon has to be compiled for NLP purposes: if we manage to reduce the types of information that a dictionary should contain, without giving up consistency and fine-grainedness, we will reduce access time and memory requirements tremendously. From a purely linguistic point of view, the possibility of generalizing is of course just as appealing. The question which arises is therefore whether or not we can come up with a rule which would say:

*If a verb expresses a change of colour, then this verb is ergative, i.e. displays the causative/inchoative alternation.*

This would mean that whenever there is a verb which is associated with a particular colour, this verb is ergative. This type of generalization would enable us to have more economic lexical entries since we would no longer need to code this explicitly. Unfortunately, our analysis of the verbs which denote a change of colour shows that this is not the case, at least for the verbs «tincture», «tinge» and «tint» which are considered by all dictionaries as transitive verbs in all their senses. Dictionaries disagree as to the ergative property of «bronze» (only transitive in *LDOCE* and clearly ergative in the *Oxford Advanced Learner's Dictionary of Current English*, 1980). The same disagreement applies to «blue» which is clearly ergative only in the *Random House Dictionary* (it is transitive in all other dictionaries). This shows that we have to be extremely careful before generalizing and that we can do so only when we have succeeded in reaching an agreement, which proves very difficult with borderline cases. If generalization is impossible, the only solution is to view ergativity as an idiosyncratic type of information to be explicitly coded at definition level.

### Colour verbs and subject field information

Besides the various types of morpho-syntactic information, the magnetic tape of the *Longman Dictionary of Contemporary English* also contains subject field information in the form of 2-byte codes à la Merriam-Webster which indicate the domain in which a particular sense of a word is used (e.g. BZ = Business; MD = Medicine...). The classification includes a code for colour terms, viz. CO. Since the dictionary is organized in database format, it is very easy to query all items which bear this code together with a part of speech denoting a verb (*v*, *v prep*, *v adv* or *v prep adv*). The aim is of course to enlarge our set of verbs expressing a change of colour and to subsequently analyze them in terms of their ergative behaviour. Unfortunately, only 7 verbs are assigned the code CO which shows that this type of information, although extremely useful with respect to word sense assignment and disambiguation, is far from being used consistently. The seven verbs in question are:

COLOUR (1), COLOUR (2), COLOUR IN (1), COME OUT (8), TAKE (32), TINGE (1) and YELLOW (1).

The only verb which had not yet been discovered is TAKE (32) (as in: «the colour took»). However, this verb can only be used intransitively and does not display the causative/inchoative alternation.

### Sound terms

We now turn to the second part of our research. As we have already mentioned, another sub-set of ergative verbs includes verbs which express the production of a sound or a noise. This is best illustrated by the following example (from *COBUILD*):

*... the sound of the bells clanging...  
she was methodically clanging the brass bells*

CLANG is defined in *LDOCE* as:

[T1,10]: to (cause to) make a loud ringing sound such as when a metal is struck.

We see that such a verb has a well-defined thematic structure: it involves an AGENT (usually the person who is responsible for the production of the sound) and a PATIENT (the concrete object which produces this sound). If we postulate a hierarchy of semantic roles whereby priority is given to the agent, the subject slot will be filled by the higher-level semantic role in the sentence (further details on the precedence scale of semantic roles can be found in Allerton, 1982). Since the realization of the agent is optional in this case, the patient will be the surface subject when the agent is not mentioned. Conversely, when the two semantic roles are present in the sentence, the agent is the surface subject and the patient appears under the form of a direct object. Our preliminary list of ergative verbs contained 35 verbs expressing the production of a sound. As can be seen below, the definition patterns include a combination of verbs denoting the concept of MAKING and words (nouns, adjectives or adverbs) denoting a sound:

anglicize (1)	become + sound
bang (2)	knock, beat, push + noise
blow (5)	sound
boom out (1)	come out + sound
bounce (2)	move + noisily
chime (1)	make + sounds
clack (1)	make + sound
clang (1)	make + loud + sound
clank (1)	make + loud + sound
clash (3)	make + loud + noise
clatter (1)	move + sounds
click (1)	make + sound
clink (1)	make + sound
crack (1)	make + sound
crackle (1)	make + sounds
crash (1)	have + noisy



crash (2)	fall, strike + noisily
echo (1)	come back + ECHO (=sound)
fade in (1)	mix + sound
fade out (1)	disappear + sound
honk (1)	make + HONK (=sound)
hoot (1)	make + hoot (=sound)
jangle (1)	make + sound
jingle (1)	sound + jingle (=sound)
pop (1)	make + sound
rattle (1)	make + noises
reecho (1)	be repeated + ECHO (=sound)
rustle (1)	make + sounds
screech (2)	make + noise
splash (2)	fall, strike, move + noisily
swish (1)	cut + making + noise
ting (1)	make + sound
tinkle (1)	make + sounds
toot (2)	produce + sound
twang (1)	make + TWANG (=sound)

As can be seen, the genus term is quite often «make» or «produce», which may be considered as surface realizations of a semantic primitive expressing production. It should however be noted that this creation process is sometimes implicit when the stress is put on the way in which the action is performed (see «bang» or «splash»). The common characteristic of these verbs is the fact that they all include words which refer to the object of the creation, namely a sound. The various items are the following: «sound», «sounds», «loud», «noise», «noises», «noisy», «noisily». We therefore extracted from our *LDOCE* database all the verbs which contained at least one of the following strings in their definition field: SOUND, LOUD or NOIS (the latter string accounting for «noise(s)», «noisy» or «noisily»). We managed to retrieve a list of 273 verbs which denote the production of a sound. We then submitted this list to a careful analysis in order to test their possible ergativity. Ideally, we should have been able to postulate a rule saying that all verbs which express the production of a sound are ergative (see our arguments for a maximal generalization). Unfortunately, this is far from being the case and we had to classify our verbs into further sub-classes. We present here the main types of verbs which fall under the more general heading of production-of-sound verbs:

- a) verbs with AGENT - PATIENT (the ergative verbs listed above);
- b) verbs with AGENT only;
- c) verbs with AGENT - EFFECTED OBJECT.

The second class comprises 47 verbs whose possible subject is mentioned within the definition itself: those verbs have a slot for an animate subject (animal and/or human). The predicate-argument structure will then contain reference to an AGENT only which appears under the form of a surface subject. Since there is no patient, the verb can only be inchoative. These verbs actually refer to the way an animal utters a sound. Since you cannot cause an animal to produce a sound in the same way as you

can cause a metal tool to, say, «cling» or «bang». these verbs will not display any causative reading. The analysis of that class of verbs enables us to semi-automatically derive the semantic relationship which links an animate being to the verb referring to the sound or cry it produces. We reproduce a sample list of these verbs:

baa	(sheep)
bark	(dog)
bellow	(bull)
bleat	(calf)
cackle	(hen)
caterwaul	(cat)
chirp	(insects)
coo	(dove)
croak	(frog)
drone	(bees)
gobble	(turkey)
quack	(duck)
trumpet	(elephant)
...	

It should be noted that the thesauric relationship which links the two columns above is actually one of the various lexical functions used by Igor Mel'cuk in his *Explanatory Combinatorial Dictionary*: in the present case, the standard basic lexical function is «Son» (to emit characteristic sounds): e.g. Son (dog) = bark. (For more details on the structure and the possible uses of the *ECD*, we refer the reader to Mel'cuk and Zholkovsky, 1988).

When looking at the list of sound verbs which are not ergative and whose predicate-argument structure involves more than just a patient, we find another 187 verbs including, among others, verbs such as «hiccup», «sing», «jar», «lisp», etc. All these verbs actually involve an AGENT and an EFFECTED OBJECT, i.e. an object entity which comes into being as a result of the verbal activity (for more details on the distinction between effected object and patient, see Allerton, 1982). This explains why many sound verbs are coded as transitive, which accounts for sentences such as:

- The radio blared out the news (BLARE OUT (1) [T1])
- \* The news blared out
- She whispered a few words (WHISPER (1) [T1,I0])
- \* A few words whispered

These verbs typically have an inchoative reading, which rules out the causative one since no patient is involved. These considerations enable us to postulate a rule which makes it possible to generalize and therefore be more economic with respect to the amount of information assigned to lexical items:

*If a verb expresses the production of a sound or noise and if the predicate-argument structure of this verb involves the semantic roles of agent and patient, then this verb is ergative.*

The retrieval of verbs which express the production of a sound again enabled us to locate a few items for which the causative/inchoative alternation had not been detected in our previous study. The ten verbs in question are given below (together with the keyword in the definition):

flap (1,2)	(... making a noise - ... making a noise)
frizzle (1,2)	(... make ... noises)
peal (1,2)	(... to sound loudly)
play (14,17)	(... produce sound - ... reproduce sounds)
rhyme (1,2)	(... end with the same sound)
ring (1,2)	(... to sound - ... to sound)
scrunch (1,2)	(... make a sound)
slam (1)	(... shut loudly)
snap (4)	(... move ... to cause such a short sound)
sound (2,3)	(... make a sound - ... make a sound)

Eight of these verbs had not been detected because the Longman lexicographers had actually opted for the splitting strategy and considered the transitive and intransitive usages as two different readings altogether. PEAL clearly illustrates this problem:

PEAL (1): [I0(OUT)]: (esp. of bells) to ring out or sound loudly (and continually) (2): [T1]: to cause (bells) to ring out

The splitting strategy exemplified above of course enables the lexicographer to specify the patient argument (bells) within the definition itself. This is obviously an advantage but we must admit that the lumping strategy is a space-saving technique which makes it possible to keep track of what the two definitions have in common.

As to SLAM and SNAP, those verbs had not been retrieved because they make use of an ergative verb as genus term (respectively «shut» and «move»).

## Conclusions

We have tried to show that the computer proves to be an efficient tool to investigate the structure of the lexicon and that it provides diversified access to move further towards the ideal dictionary. Our onomasiological approach proves to be an excellent way of discovering the relationships that link lexical items. The lexicon is no longer seen exclusively as a repository of irregularities and idiosyncratic information, but must be viewed as a network whose nodes (the actual words) are closely interrelated in a web-like fashion. The arrows that connect those nodes are labels for lexico-semantic relations and this structure can be investigated by exploiting the regularities which underlie the lexicon: morphological, syntactic and semantic phenomena can be studied with a view to producing, improving and enhancing the quality of lexical databases usable in natural language processing.

## References

- AHLISWEDE, T., EVENS, M. (1988): «Generating a Relational Lexicon from a Machine-Readable Dictionary». In Ison, R. (ed.), *International Journal of Lexicography*, Vol. 1, Number 3, Oxford University Press, Oxford.
- ALLERTON, D. (1982): *Valency and the English Verb*. Academic Press.
- ALSHAWI, H. (1987): «Processing Dictionary Definitions with Phrasal Pattern Hierarchies». *Computational Linguistics*, Vol. 13, 3-4.
- ATKINS, B., KEGL, J., LEVIN, B. (1987): «Explicit and Implicit Information in Dictionaries». Lexicon Project Working Papers 12, Center for Cognitive Science, MIT, Cambridge, Massachusetts.
- ATKINS, B., KEGL, J., LEVIN, B. (1988): «Anatomy of a Verb Entry: From Linguistic Theory to Lexicographic Practice». In Ison, R. (ed.), *International Journal of Lexicography*, Vol. 1, Number 2, pp. 84-127, Oxford University Press, Oxford.
- BOGURAËV, B., BRISCOE, T. (1987): «Large Lexicons for Natural Language Processing: Utilising the Grammar Coding System of LDOCE». *Computational Linguistics*, Vol. 13, 3-4.
- BOGURAËV, B., BRISCOE, T. (1989): *Computational Lexicography for Natural Language Processing*. Longman Group Ltd., Harlow.
- BYRD, R., CALZOLARI, N., CHODOROW, M., KLAVANS, J., NEFF, M., RIZK, O. (1987): «Tools and Methods for Computational Lexicology». *Computational Linguistics*, Vol. 13, 3-4.
- CALZOLARI, N. (1984): «Detecting Patterns in a Lexical Data Base». *Proceedings of COLING 84*, ACL, Stanford, California.
- CALZOLARI, N. (1988a): «Structure and access in an automated lexicon and related issues». In Walker, D., Zampolli, A., Calzolari, N. (eds.) *Automating the Lexicon: Research and Practice in a Multilingual Environment*. Cambridge University Press, Cambridge.
- CALZOLARI, N. (1988b): «The Dictionary and the Thesaurus Can Be Combined». In Evens, M. (ed.) *Relational Models of the Lexicon*. CUP, Cambridge.
- EVENS, M., LITOWITZ, B., MARKOWITZ, J., SMITH, R., WERNER, O. (1980): *Lexical-Semantic Relations: A Comparative Survey*. Edmonton, Linguistic Research, Inc.
- EVENS, M. (1988) (ed.): *Relational Models of the Lexicon: Representing Knowledge in Semantic Networks*. Cambridge University Press, Cambridge.
- FONTENELLE, Th., VANANDROYE, J. (1989): «Retrieving Ergative Verbs from a Lexical Data Base». In R. Bailey (ed.), *Dictionaries: Journal of the Dictionary Society of North America*, Vol. 11, pp. 11-39.
- HORNBY, A. S. (1980) (ed.): *Oxford Advanced Learner's Dictionary of Current English* (3rd edition). Oxford University Press, Oxford.
- LEHRER, A. (1974): *Semantic Fields and Lexical Structure*. Amsterdam-London, North-Holland.
- LEVIN, B. (1988): «Approaches to lexical semantic representation». In Walker, D., Zampolli, A., Calzolari, N. (eds.), *Automating the Lexicon: Research and Practice in a Multilingual Environment*. Cambridge University Press, Cambridge.
- LYONS, J. (1969): *Introduction to Theoretical Linguistics*. Cambridge University Press, Cambridge.
- MEL'CUK, I., ZHOLKOVSKY, A. (1988): «The Explanatory Combinatorial Dictionary». In Evens, M. (ed.), *Relational Models of the Lexicon*, CUP, Cambridge.
- MICHELIS, A. (1982): *Exploiting a Large Dictionary Data Base*. Ph. D. Thesis, University of Liège, Liège, Belgium.
- MICHELIS, A., NOEL, J. (1982): «Approaches to Thesaurus Production». *Proceedings of COLING 82*, ACL, Prague.
- PROCTER, P. (ed.) (1978): *Longman Dictionary of Contemporary English*. Longman Group Ltd., Harlow.
- The Random House Dictionary of the English Language* (1983): Second edition, unabridged. Random House, New York.
- SINCLAIR, J. (ed.) (1987): *Collins COBUILD English Language Dictionary*. William Collins Sons and Co. Ltd., London.

WILKS, Y., FASS, D., GUO, C., McDONALD, J., PLATE, T., SLATOR, B. (1989): «A tractable machine dictionary as a resource for computational semantics». In Boguraev, B., and Briscoe, T. (eds.) *Computational Lexicography for Natural Language Processing*. Longman.

## APPENDIX

Verbs expressing the production of a sound or noise:

Class A: Ergative verbs (verbs with AGENT + PATIENT)

anglicize(1)	click(1)	frizzle(1)	rattle(1)	snap(4)
bang(2)	clink(1)	frizzle(2)	reecho(1)	sound(2)
blow(5)	crack(1)	honk(1)	rhyme(1)	sound(3)
boom out(1)	crackle(1)	hoot(1)	rhyme(2)	splash(2)
bounce(2)	crash(1)	jangle(1)	ring(1)	swish(1)
chime(1)	crash(2)	jingle(1)	ring(2)	ting(1)
clack(1)	echo(1)	peal(1)	rustle(1)	tinkle(1)
clang(1)	fade in(1)	peal(2)	screech(2)	toot(2)
clank(1)	fade out(1)	play(14)	scrunch(2)	twang(1)
clash(3)	flap(1)	play(17)	slam(1)	
clatter (1)	flap (2)	pop (1)	snap (3)	

Class B: verbs expressing the typical sound/cry of an animate being (verbs with AGENT only, retrieved from the definition)

baa	(sheep)	drone	(bees)
baa	(lamb)	gobble	(turkey)
bark	(dog)	grizzle	(children)
bay	(dog)	growl	(animals)
belch	(person)	grunt	(animals)
bellow	(bull)	grunt	(human beings)
bleat	(sheep)	gurgle	(babies)
bleat	(goat)	hum	(people)
bleat	(calf)	mew	(gull)
bray	(donkey)	miaow	(cat)
cackle	(hen)	moo	(cow)
call	(animal)	neigh	(horse)
caterwaul	(cat)	oink	(pig)
cavort	(person)	pule	(baby)
caw	(crow)	quack	(duck)
caw	(bird)	roar	(child)
champ	(horse)	scoop	(singer)
chomp	(horse)	snarl	(animal)
chatter	(animals)	snicker	(horse)
chatter	(birds)	snuff	(animals)
cheep	(birds)	squawk	(birds)
chirp	(birds)	stridulate	(insects)

chirp	(insects)	strike	(person)
clop	(horses' feet)	trumpet	(elephant)
coo	(dove)	tweet	(bird)
coo	(pigeon)	twitter	(bird)
croak	(frog)	wheez	(people)
crow	(cock)	whimper	(person)
crow	(baby)	whinny	(horse)
cry	(animals)	yap	(dog)
cry	(birds)	yowl	(animal)

## Class C: verbs with an AGENT and an optional EFFECTED OBJECT

air(3)	emit(1)	ping(1)	smack(1)
amplify(2)	escape(5)	pink(1)	smack(3)
aspirate(1)	fizz(1)	pitch(6)	sniff(1)
aspirate(2)	gasp(2)	pound(2)	snore(1)
babble(5)	give back(2)	pound(4)	snort(1)
bang(3)	give out(6)	pound(6)	snort(2)
bark(2)	glide(1)	project(1)	sob(1)
beat out(1)	go(19)	pronounce(1)	sound(4)
blare(1)	gobble(1)	pulse(1)	sound(5)
blare out(1)	go off(1)	purl(1)	speak(7)
blubber(1)	grate(2)	purr(1)	speechify(1)
bluster(1)	grind(3)	rag(1)	splutter(2)
boo(2)	growl(2)	rail(1)	sputter(2)
boohoo(1)	growl(3)	raise(10)	squabble(1)
boom(1)	grunt(3)	raise(11)	squall(1)
bubble(2)	hear(1)	rasp(3)	squeak(1)
burble(1)	hiccup(2)	rattle(2)	squeal(1)
burst in on(1)	hiss(1)	receive(5)	squelch(2)
bustle(1)	hiss off(1)	record(2)	stammer(1)
bustle with(1)	hoot down(1)	reflect(1)	stutter(1)
cackle(2)	howl down(1)	resonate(1)	swish(2)
call(8)	huff(1)	resonate(2)	thrum(1)
chime(2)	insulate(1)	resound(1)	thump(2)
chirp(2)	jam(5)	resound(2)	thunder(2)
chug(1)	jangle(2)	reverberate(1)	tick(1)
clap(1)	jar(1)	ring out(1)	toot(1)
clump(1)	keen(2)	ring with(1)	utter(1)
collocate(1)	knock(1)	ripple(3)	velarize(1)
cough(1)	knock(5)	roll(8)	vocalize(1)
cough(3)	lap(2)	romp(1)	vocalize(3)
crash(3)	laugh(1)	row(1)	voice(2)
crash(4)	lilt(1)	rumble(1)	wail(1)
creak(1)	lisp(1)	rumble(2)	wake(6)
create(3)	listen out(1)	rustle(2)	whack(1)
crunch(1)	melt(4)	say(1)	wheeze(1)
crunch(2)	moan(1)	scratch(2)	whine(1)

cry(1)	modulate(1)	scream(4)	whirr(1)
cry(2)	muffle(1)	shriek(1)	whisper(1)
damp(1)	munch(1)	sigh(1)	whisper(2)
demagnetize(1)	munch(1)	sigh(2)	whistle(1)
die away(1)	murmur(1)	silence(1)	whistle(4)
din(1)	mute(1)	sing(1)	whiz(1)
disparage(1)	patter(1)	sing(3)	wind(1)
drown(6)	patter(2)	sing(4)	yammer(2)
drum(3)	pause on(1)	sizzle(1)	yap(2)
dub(1)	penetrate(2)	slur(3)	zip(2)
elide(1)	pierce(4)	slurp(1)	