IDAREX: Formal Description of Multi-Word Lexemes with Regular Expressions*

Frédérique Segond, Giuseppe Valetto
Rank Xerox Research Centre,
6 chemin de Maupertuis,
38240 Meylan, France
frederique.segond@xerox.fr
valetto@xerox.fr

Elisabeth Breidt
Seminar für Sprachwissenschaft,
Universität Tübingen,
Wilhelmstr. 113,
D-72074 Tübingen, Germany
breidt@sfs.nphil.uni-tuebingen.de

13 December 1995

Abstract

Most multi-word lexemes (MWLs) allow certain types of variation. This has to be taken into account for their description and their recognition in texts. We suggest to describe their syntactic restrictions and their idiosyncratic peculiarities with local grammar rules, which at the same time express in a general way regularities valid for a whole class of MWLs. The local grammars can be written in a very convenient and compact way as regular expressions in the formalism IDAREX which uses a two-level morphology. IDAREX allows the definition of various types of variables, and to mix canonical and inflected word forms in the regular expressions.

1 Introduction

Most texts are rich in multi-word expressions that cannot be properly translated – or even understood – if they are not recognized as complex lexical units. We

*Submitted to COLING 1996, Copenhagen, Denmark.
This work is part of the Locolex/Compass project (LRE research project 62-080)whose main aim is the development of an on-line foreign language comprehension tool. It presupposes work done by other members of the Locolex/Compass team, especially Lauri Karttunen and Pasi Tapanainen. Approximately 10,000 English and 4,800 German MWLs have been described so far in the project. A description of Locolex is given in Bauer, Segond and Zaenen (1995). The dictionaries used in the project are the Oxford-Hachette English-French dictionary (1994) and the HarperCollins German-English dictionary (1991). We are grateful to the publishers to give us the right to use them for research purposes. We thank Annie Zaenen, Ted Briscoe, and Irene Maxwell for their comments on an earlier draft of this paper.

call such expressions *multi-word lexemes (MWL)*. These range from idioms (*to rack one's brains over sth*), over phrasal verbs (*to come up with*), lexical and grammatical collocations (*to make love, with regard to*) to compounds (*online dictionary*).

It is well known (see for instance, Nunberg et al. (1994), Gross (1975), Brundage et al. (1992)) that most MWLs cannot be treated like completely fixed patterns, since they may undergo some variation. However, only a *subset* of the range of variation allowed by the rules of the general syntax is valid: outside that subset, the expression loses its idiomatic meaning, either reverting to its literal meaning or losing any significance altogether. In certain cases, MWLs can even contradict normal syntactic rules, as with the German expression *von Haus aus* (originally), where general rules would require an article. Such MWLs have to be recognized before parsing because otherwise the parser would fail.

Therefore, in order to recognize the full range of occurrences of MWLs, it is necessary to develop a method that describes their syntactic restrictions and their idiosyncratic peculiarities on a case-by-case basis.

NLP treatments of MWLs have been proposed in e.g. Abeillé and Schabes (1989) in the framework of lexicalised TAGs, Erbach (1992), and Copestake (1994) in HPSG, Van der Linden (1993) in CG. These proposals discuss ways in which MWLs should be treated in so-called high level grammar formalisms. Within the finite state approach to NL parsing, it has been proposed to represent MWLs by 'local grammars' (e.g. Roche (1993), Maurel (1993), and Silberztein (1993)). Local grammars as discussed in these contributions are finite state networks that are matched against (parts of) a NL string. Informally speaking, if the string corresponds to one traversal of the network, it is recognised by the local grammar represented by that network.

In the proposal made by the 'Parisian school' (i.e. the researchers working on finite state grammars at Paris VII some of which are mentioned above) the networks are created as such. Under the approach developed at RXRC they are compiled from descriptions which take the form of regular expressions (REs). The compiler written by Lauri Karttunen, Pasi Tapanainen and Giuseppe Valetto and described in Tapanainen (1994), allows the sequence of regular expressions to mix canonical and inflected forms.

Encoding the local grammars as REs instead of encoding them directly as networks does of course not change the expressive power of the formalism, but it conveniently abstracts the handling of MWLs from the graph manipulation level, allowing to develop and employ devices that operate on string representations and map them to the underlying finite state networks. As we will show below, this simplifies considerably the description of the different patterns of variation occuring in MWLs.

2 Variability of MWLs and Local Grammar Rules

2.1 Variability of MWLs

Certain MWLs always occur in exactly the same form, like *out of the blue* or *um Haaresbreite* (almost); these can be easily recognised with simple pattern matching techniques. However, most MWLs allow some types of variation, flexibility and/or modification[1], although they do not have the full (morpho-)syntactic potential of free word combinations.

Basically, we recognize four types of variability[2], which we enumerate below.

Morphological Variation: Particular words in the MWL can undergo certain inflections. For instance:

In *to kick the bucket*, *kick* can be inflected to any tense, number, person. In *faire le fou* (to play the fool) the determiner and the noun can appear in their feminine form, respectively *la* and *folle*. In *durchschlagender Erfolg* (sweeping success), noun and adjective can vary in case and in number; besides, comparative and superlative of the adjective are allowed.

Lexical Variations: One or more words can be substituted by other terms, e.g. near-synonyms, without changing the overall meaning of the MWL. For instance:

In *to sweep sth under the carpet*, *rug* or *mat* can be substituted for *carpet*.In *sich aufs Ohr legen* (to hit the hay), the exchangeable components have no meaning similarities at all: *hauen* (to hit) can be used instead of *legen* (to put/lay down), and *aufs Ohr* (onto the ear) can be replaced by *in die Falle* (into the trap).

Modifications: One of the MWL's constituents can be modified, while the idiomatic meaning is preserved. For instance:

In *it's someone's business* one can insert a negation and get *it's not someone's business*. In the French MWL *avoir un coup dans l'aile* (to have drunk too much) an adjective can be inserted, and the idiomatic nature of the expression is preserved, e.g. *elle a un sacré coup dans l'aile*. In *den (schönen) Schein wahren*, the adjective modification does not alter the meaning at all.

Structural Flexibility: These include phenomena like passivization, topicalization, scrambling, raising constructions etc. For instance:

The MWL *to keep tabs on* can be passivized without loosing its idiomatic meaning. Similarly, the French expression *chercher midi à quatorze heures* (to nit pick) can be topicalized: *Midi, il ne le cherchait pas à quatorze heures*. In German, standard word order variation (e.g. Uszkoreit 1987) applies to all verbal MWLs: idiom-external arguments can be scrambled in the middle field and topicalized according to general rules. Scrambling of a fixed component is mainly possible with an idiom-external prepositional argument, as in *für etw den Kopf hinhalten/den Kopf für etw hinhalten* (to take the blame for sth). In certain cases[3], even the fixed component may be topicalized, as in *den Vogel dabei hat dann Jan abgeschossen* (finally, Jan surpassed everyone).

[1]By *variation* we mean that different lexical items can occupy the same position in a MWL; by *flexibility* we understand a syntactic re-structuring of components; by *modification* we understand the addition of modifying words to the MWL.

[2]See also Fleischer (1982), Brundage et al. (1992), Engelke (1994) and Van der Linden (1993).

[3]This seems to depend on the metaphorical relation of the idiomatic and the literal meaning of the component.

3 The MWL Formalism IDAREX

We write the local grammar rules in the formalism IDAREX (**ID**ioms **A**s **R**egular **EX**pressions) which uses a two-level morphology and has been developed as part of FSC (Finite State Compiler) at Rank Xerox Research Centre[4]. We encode MWLs as regular expressions according to the following notations.

3.1 Word Level

Words can be represented at two levels: lexical level and surface level. We allow four basic descriptions for individual words:

1. :surface-form
2. :surface-form morphological-variable:
3. lexical-form morphological-variable:
4. word-class-variable

The first two cases stand for **fixed forms** of words. For instance, *pédales* in the French expression *perdre les pédales* can be coded as:

1. `:pédales`
2. `:pédales Noun:`

The third class stands for **morphologically variable forms** of words. In the previous example, the verb *perdre* can appear in any tense, number and person. Therefore it is coded as:

3. `perdre Verb:`

which expresses that we may take any verb instance of the word *perdre*. More restrictive morphological variables can also be defined, e.g. to restrict a noun to singular or plural, or an adjective to comparative form, as in

höhere Gewalt (act of God) $\neq$ *hohe/höchste Gewalt(en)* (high(est) force(s))
$\Rightarrow$ `hoch Acomp: Gewalt Nsg:`

The last class, word-class variables, stands for any word belonging to a **particular word-class**. For instance, if we define a variable `ADV` for all the adverbials or sequences of adverbials, we may finally write the French idiom under discussion as:

`perdre Verb: ADV* :les :pédales`

In general, complex variables may be defined as sequences of basic word-class variables, enabling the definition of complex syntactic categories.

Morphological and word-class variables can be seen as shortcuts for particular regular expressions which cover specific sequences of morphological and lexical information. Their definition depends on the exact form of the morphological analysis.

[4]For a more detailed description of the formalism see Karttunen and Beesley (1992), Karttunen and Yampol (1993), Tapanainen (1994), Karttunen (1995), Kempe and Karttunen (1995), and Segond and Tapanainen (1995).

3.2 Operators

In addition, we use a set of regular expression operators to combine the word descriptions. Among them are:

- *nothing* — the words are simply concatenated.
- *parentheses* () — mark an optional part of an idiom.
- *Kleene star* * — denotes that the preceding expression may appear any number of times, including null.
- *plus* + — denotes that preceding expression may appear any number of times, but has to occur at least once.
- *brackets* [] — groups an expression.
- *bar* | — separates alternative expressions.
- *semicolon* ; — finishes an expression.

Parentheses and Kleene star are used to encode possible modifications, as shown in the examples below:

(i) *prendre (souvent) le taureau par les cornes*
(to (often) take the bull by the horns)
⇒ `prendre Verb: ADV* :le :taureau :par :les :cornes`

(ii) *von (ADJ) Vorteil sein* (to be ... advantageous)
⇒ `[ :von (ADJ) :Vorteil (:zu) sein V: |...`[5]]

With the bar and brackets, lexical variations and alternations of more complex sequences, e.g. the attributive negation and word order variation in German, are encoded:

(iii) *perdre la tête/la boule/les pédales* (to lose one's mind)
≠ *perdre la tronche* (to lose one's mug)
⇒ `perdre Verb: [ :la :tête | :la :boule | :les :pédales ]`

(iv) *(k)eine (dicke) Lippe riskieren* ((not) to be brazen)
⇒ `[ [ :eine | :keine ] (:dicke) :Lippe (:zu) riskieren V: |`
`riskieren Vfin: (ADV* NPnom) ADV* [:eine|:keine] (:dicke) :Lippe]`

3.3 Syntactic Level: Macros

Morphological and word-class variables can represent variability at the level of the single word. So, variables provide an efficient abbreviation mechanism and are a essential for compact idiom definitions. However, MWLs allow additional levels of variability, which involve more complex constituents. Interesting examples: the conjugation of negated verb in English, reflexive verbs in French, word order variation in German, including topicalization, passivization and separable prefix verbs.

[5] Second part of RE for V1/V2 word order omitted.

Again, it is convenient to handle such phenomena with local grammar rules; to achieve this, **macros** accepting multiple parameters have been adopted as part of the notation. The power of these macros resides mainly in their parametrization that permits to describe with a single operator the regularities that can be observed for a whole class of MWLs, and to apply them to any instance of this class. Each macro is made up of three components: (i) its name, (ii) the set and number of parameters it accepts, and (iii) its definition or 'body', made of regular IDAREX constructs and numeric place-holders for the run-time insertion of the parameters.

A Macro can be inserted within any MWL representation by using its name and specifying, between parentheses and separated by spaces, the parameters that must be substituted to the place-holders at the moment the macro is expanded by the IDAREX compiler. It is worth mentioning that the parameters can be single words, in their lexical or surface form, variables, operators or even other macros with their own parameters. This offers a lot of flexibility and thus the possibility to represent complex phenomena in a compact way.

We want to illustrate the usefulness and expressive power of macros with standard word order variation of German verbal MWLs. Three basic orderings are possible, verb-first (V1), verb-second (V2) and verb-final (Vend), and an infinitive construction (Inf+zu).

V1: *Kommt Jan dir (dabei) in die Quere?* (does Jan cross your path (in this)?)

V2: *Jan kam Ute (dabei) in die Quere.* (Jan crossed Ute's path)

Vend: *Ute war sauer, weil Jan ihr (dabei) in die Quere kam.*
(Ute was angry because Jan had crossed her path)

Inf+zu: *Jan versucht, Ute (dabei) nicht in die Quere zu kommen.*
(Jan tries not to cross Ute's path)

Besides, topicalization of the external argument (here *Jan*) and scrambling with the subject (*Ute*) is possible in all three word orderings.

V1: *Kommt dir (dabei) Jan in die Quere?*

V2: *Ihr kam (dabei) Jan in die Quere.*

Vend: *Ute war sauer, weil ihr Jan (dabei) in die Quere kam.*

All these variations are expressed with the following macro use and definition:

```
V1ArgWO(NPdat fix3(:in :die :Quere) kommen) ;
```

`V1ArgWO` is defined as

```
[ $3 Vfin: 1ARG1($1) $2 | $2 (Vaux ( 1ArgEnd($1) )) (:zu) $3 V:]
```

The macro `V1ArgWO` covers word order variation of verbal MWLs with one external argument. It is defined in a very general way and allows different types of external argument as parameter $1, i.e. a dative or accusative object (NPdat, NPakk), or a

PP. For verbal MWLs having two external arguments (e.g. *jdm etw in Rechnung stellen* (to charge sb for sth)), a separate macro has to be used. \$3 is instantiated with the verb, and \$2 with the remaining fixed elements of the MWL. A parameter can only be a single item, but the number of fixed words varies for each verbal MWL, so 'assist' macros `fix(i)` are defined for different lengths. For example, `fix3` is defined as `$1 $2 $3`. The first part of the disjunction covers V1 and V2 order, the second part matches verb-final order and the infinitival construction. `1ARG1` and `1ArgEnd` are further macros taking care of standard and scrambled ordering of subject and external argument for V1/V2 and Vend ordering, respectively.

1ARG1 is defined as

`[ (ADV* NPnom) ADV* $1 | (ADV* $1) ADV* NPnom ] ADV*`

1ArgEnd is defined as

`ADV* [ NPnom ADV* $1 | $1 ADV* NPnom ] ADV*`

In both definitions, \$1 is instantiated with the variable for the subcategorized complement, i.e. NPakk, NPdat, or with a macro for PPs[6]. In `1ARG1`, the first component in each part of the disjunction is made optional because in V2 order, this component is moved to topic position to the left of the verb.

In addition, the second part of `V1ArgWO` covers topicalization of the fixed MWL component (*in die Quere*). Note that this is only possible for certain verbal MWLs and in certain contexts.

Top: *In die Quere hatte Jan ihr sicher nicht kommen wollen.*
(certainly, Jan had not wanted to cross her path)

The above example for the use of `V1ArgWO` leads to the following RE when the macro is expanded:

`V1ArgWO(NPdat fix3(:in :die :Quere) kommen) ;` expands to

```
[kommen Vfin: [(ADV* NPnom) ADV* NPdat | (ADV* NPdat) ADV* NPnom]
ADV* :in :die :Quere | :in :die :Quere (Vaux ([ NPnom ADV* NPdat
| NPdat ADV* NPnom ] ADV*)) (:zu) kommen V:]
```

For German, similar macros are defined for verbal MWLs where the fixed component can be scrambled with an idiom-external PP argument, for MWLs with a reflexive verb, for prefix verbs in their separated form, and for MWLs with a separable prefix verb. In French, macros describe for example the verb complex for reflexive verbs and constructions with the pronoun *en*.

4 Discussion

The identification of MWLs is essential for any kind of natural language processing based on lexical information, ranging from intelligent concordancing or indexing over automatic dictionary look-up to machine translation. As MWLs are not simply

[6]The PP macro can be used to express the generally possible alternation of a PP with a pronominal adverb, e.g. *mit etw* vs. *damit*.

lexical patterns but also show syntactic behaviour, their existence pleads for a tighter interaction between syntax and lexicon than has often been assumed. We have illustrated how this can be achieved using a finite-state approach and local grammar rules formalized as regular expressions.

Finite State techniques are known to be unable to represent all the dependencies found in natural language. They are, however, capable of representing a great number of phenomena in an efficient way. Their use is hampered by the unwieldiness in notation that these techniques often lead to. The development of adequate notational devices and corresponding compilers is therefore essential. We see three advantages of the system that we have described: its compactness, its flexibility and its modularity. As we saw, we can define generic variables, such as "any adverb" (ADV) or much more specific morphological variables, such as "only verbs in the third person singular" (Vsg3). This mechanism allows to define as many levels of abstraction as necessary for encoding MWLs, thus greatly relieving lexicographers from the burden of explicitly listing all the possible forms. Moreover, macros provide a means to formulate generalizations about patterns that can occur for a whole class of MWLs. Finally, the ability to state patterns on both the lexical and the surface level enables us to express generalisations on either level. Therefore, it is very easy to restrict the use of an MWL constituent to any of its possible inflections, by merely specifying it as a surface form and without bothering to consider and rule out all the features of the lexical side. The compactness and flexibility as illustrated follow from the possibility to encode the network as regular expressions. These advantages are, as far as we know, specific to our approach.

Another interesting property of the finite state approach in general is its modularity. This is an advantage our approach shares with the Parisian notion of local grammars. The local grammars we propose encode MWL restrictions to the general syntax, but they are not meant to replace the general syntax. For instance, in a local grammar rule such as

> prendre V: (POSCLITIC) ADV * (NP) :sous POSS :bonnet ;
> (*prendre qch sous son bonnet* – to make something one's concern)

the inserted NP and adverbial constituents are external to the MWL and as such should be handled by the general syntax. The general syntax of an NP can then again be a local grammar as for instance proposed in Silberztein (1993) where all linguistic phenomena are treated as local grammars. But it need not to be used in this way. In fact, contrary to the Parisian philosophy, we intend to investigate the integration of local FS grammars with other more powerful grammar formalisms.

The local grammar rules we write cover at most sentence-length patterns. They are formulated as generally as possible, allowing for overgeneration. Although more specific and restrictive rules could be written, this is unnecessary as long as the input is grammatical. In this respect, our matching approach differs from high level parsing. In high level parsing the goal is to constrain the grammar to match all and nothing but grammatical input. Here, the most absurd matches could be obtained if the input were not grammatical. E.g. the pattern defined in section 3.3 will also match the following ungrammatical strings of German:

Kamen oft dem Mann in die Quere.
Das Kind kam aber dem Mann nicht das Wetter in die Quere.
In die Quere wollten ihm unter Umstaenden ich in aller Eile zu kommen.

In fact, even grammatical input can sometimes be matched wrongly. To take a very simple example, the pattern '`:en :plus`' is associated with the translation 'extra, more, too many' e.g. in *cela vous coûtera 3 francs en plus* (that will cost you 3 francs more) but it can also be part of a longer expression '`:de :plus :en :plus`' as in *la vie est de plus en plus chère* (life is more and more expensive). For this reason, applications employing our system to match their input against IDAREX representations of MWLs should always take care to select the longest among multiple matches. It is not sure, however, that the relation between two matching MWLs is always that of simple inclusion of the one in the other.

The technique has its limits: productive variations of MWLs cannot be systematically captured with our approach because such variations are by their very nature unforeseeable. Examples of this are ad-hoc compound formations or combinations of metaphors and idioms, as in

das bißchen Kopf, das sie noch haben, zerbrechen sie sich mit ... (from Fleischer 1982)
(with that little intelligence they still have they rack their brains over ...)
← *sich den Kopf zerbrechen* (to rack one's brains)
+ *Köpfchen haben/etwas im Kopf haben* (to have brains)

While it is possible to extend the technique so that the system matches also the approximate *das bißchen Kopf haben* and *sich das bißchen Kopf zerbrechen* to the two expressions involved, it will continue to fail on sentences like the one used in last year's French presidential campaign where the candidate of the extreme right dismissed his more to the left competitors by saying *c'est bonnet rouge et rouge bonnet* (it is bonnet red and red bonnet). The standard expression is *c'est bonnet blanc et blanc bonnet* (lit. It is bonnet white and white bonnet), meaning 'it's six of one and half a dozen of the other'. The change here is prompted by the association of *red* with 'to the political left' but, of course, the exploitation of this type of information is not only beyond the scope of the approach that we have presented here but also beyond that of more powerful mechanisms.

References

[Abeillé and Schabes1989] Anne Abeillé and Yves Schabes. 1989. Parsing idioms in lexicalized TAGs. In *Proceedings of the 4th EACL*, Manchester, UK.

[Bauer, Segond and Zaenen1995] Daniel Bauer, Frédérique Segond, and Annie Zaenen. 1995. LOCOLEX: the translation rolls off your tongue. In *Proceedings of the ACH-ALLC conference*, pages 6–8, Santa Barbara, California.

[Brundage et al. 1992] Jennifer Brundage, Maren Kresse, Ulrike Schwall, and Angelika Storrer. 1992. Multiword Lexemes: A Monolingual and Contrastive Typology for NLP and MT. *IWBS Report 232*, IBM TR-80.92-029, IBM Deutschland GmbH, Institut f"ur Wissensbasierte Systeme, Heidelberg, September.

[Copestake1994] Anne Copestake. 1994. Idioms in general and in HPSG. Presentation given at the Workshop 'The Future of the Dictionary', Uriage-Les-Bains, France, September.

[Engelke1994] Sabine Engelke. 1994. *Eigenschaften von Phraseolexemen: Eine Untersuchung zur syntaktischen Variabilität und internen Modifizierbarkeit von somatischen verbalen Phraseolexemen.* Masters Thesis, Universität Tübingen, Germany, April.

[Erbach1992] Gregor Erbach. 1992. Head-Driven Lexical Representation of Idioms in HPSG. In M. Everaert et al., editors, *Proceedings of the International Conference on Idioms*, Tilburg, NL, September.

[Fleischer1982] Wolfgang Fleischer. 1982. *Phraseologie der deutschen Gegenwartssprache.* VEB Bibliographisches Institut, Leipzig, Germany.

[Gross1975] Maurice Gross. 1975. *Mthodes en syntaxe.* Hermann, Paris.

[HarperCollins1991] HarperCollins and Klett. 1991. *German-English dictionary.* P. Terell, V. Schnorr, W. V. A. Morris, R. Breitsprecher, editors. 2nd Edition. HarperCollins Publishers, Glasgow, UK.

[Karttunen and Beesley1992] Lauri Karttunen and Kenneth Beesley. 1992. Two-level rule compiler. *Technical Report ISTL-92-2*, Xerox Palo Alto Research Center, California.

[Karttunen and Yampol1993] Lauri Karttunen and Todd Yampol. 1993. Interactive Finite-State Calculus. *Technical Report ISTL-NLTT-1993-04-01*, Xerox Palo Alto Research Center, California.

[Karttunen1995] Lauri Karttunen. 1995. The Replace Operator. *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL-95)*, Boston, MA.

[Kempe and Karttunen1995] Andreas Kempe and Lauri Karttunen. 1995. The Parallel Replacement Operation in Finite State Calculus. *Technical Report*, Rank Xerox Research Centre, Grenoble, France.

[Lenerz1977] Jürgen Lenerz. 1977. *Zur Abfolge nominaler Satzglieder im Deutschen.* Studien zur deutschen Grammatik 5, Gunter Narr, Tübingen.

[van der Linden1993] Erik-Jan van der Linden. 1993. *A Categorial Computational Theory of Idioms.* OTS Dissertation Series, Utrecht, NL.

[Maurel1993] Denis Maurel. 1993. Passage d'un automate avec tables d'acceptablilité à un automate lexical. In *Actes du colloque Informatique et langue naturelle*, pages 269–279, Nantes, France.

[Nunberg, Wasow and Sag1994] Geoffrey Nunberg, Thomas Wasow, and Ivan Sag. 1994. Idioms. *Language*, 70/3:491–538.

[Oxford-Hachette1994] Oxford-Hachette. 1994. *English-French Dictionary*. Oxford University Press, Oxford, UK.

[Roche1993] E. Roche. 1993. *Analyse syntaxique transformationnelle du francais par transducteurs et lexique-grammaire*. Thèse de doctorat, Université Paris 7.

[Segond and Tapanainen1995] Frédérique Segond and Tapanainen Pasi. 1995. Using a finite-state based formalism to identify and generate multiword expressions. *Technical Report MLTT-019*, Rank Xerox Research Centre, Grenoble, France, July.

[Silberztein1993] Max Silberztein. 1993. *Dictionnaires électroniques et analyse automatique de textes – Le système INTEX*. Masson, Paris, France.

[Tapanainen1994] Pasi Tapanainen. 1994. RXRC Finite-State rule Compiler. *Technical Report MLTT-020*, Rank Xerox Research Centre, Grenoble, France.