

Compositional translation

M. T. Rosetta (pseud.)

Dordrecht: Kluwer Academic Publishers (The Kluwer international series in engineering and computer science; Natural language processing and machine translation, edited by Jaime Carbonell), 1994, xviii + 478 pp. Hardbound, ISBN 0-7923-9462-3, \$88.00, £64.00, Dfl 180.00

Reviewed by
Bonnie J. Dorr
University of Maryland

This book describes the theoretical underpinnings and results of Rosetta, a machine translation (MT) project that started at Philips Research Laboratory in the early 1980's; the book focuses on research carried out between 1985 and 1992. While the project was a collective enterprise among a large number of people (as the pen name indicates), the principal authors were Lisette Appelo, Theo Janssen, Franciska de Jong, and Jan Landsbergen. The book provides a coherent distillation of dozens of publications; in particular, the work described therein has served as the basis of at least two doctoral dissertations (Janssen 1986; Appelo 1993), as well as numerous technical reports, conference papers, and journal articles. In addition to providing a novel framework for examining issues in MT, the book covers several topics that are tied together within a compositional, interlingual approach based on Montague grammar. Thus, it serves as a valuable resource for researchers in MT and computational linguistics; with some work, it could also serve as a course textbook.¹

The main thesis of the book is that MT can be a good carrier for linguistic and computational research; thus, the emphasis is on modeling linguistic knowledge involved in the translation of natural languages (English, Dutch, and Spanish). The two main principles threaded throughout the book are:

- *Principle of Compositionality*: Two expressions are each other's translation if they are built up from parts which are each other's translation, by means of translation-equivalent rules.
- *Principle of Isomorphism*: Two sentences are considered translations of each other if they have the same semantic derivation trees (hence corresponding syntactic derivation trees).

Within this framework there is a close relation between syntax and semantics: the meaning of an expression is a function of the meaning of its parts, which, in turn are defined by syntax. The translation relation corresponds to the "tuning" of grammars

¹ For example, it might be possible to include some programming problems at the end of each chapter that would allow the student to incrementally build a mini-translation system between two languages using the principles of the Rosetta approach. An additional instructional tool would be the provision of software for use in the development of such a system.

across language pairs, i.e., every lexical item and rule in the source language (SL) has a corresponding lexical item and rule in the target language (TL). The challenge of this book is to demonstrate that, by adhering as closely as possible to the principles of compositionality and isomorphism, it is possible to master a wide range of linguistic problems and to produce meaning-preserving translations.

1. Outline of the book

The book is lengthy and dense with details of the approach; however, the material is clearly presented with well-explained examples in all three languages.² In addition to the introduction (chapter 1) and conclusion (chapter 21), there are five parts to the book, each of which will be described, in turn.

Part I (chapters 2–5): *The Method*. This part of the book spells out the two main principles given above and describes the nature of the translation relation between two languages. The notion of M(ontague)-grammars is introduced: these are grammars consisting of powerful compositional rules (M-rules) that operate on constituent trees (S-trees) capable of representing various kinds of ambiguities. The translation process is described as a system of M-grammars with three basic components (morphology, syntax, and semantics), each of which is *reversible*. Chapters 4 and 5 lay the foundation for the rest of the book. The main point is that syntactic operations within the M-grammar framework are meaning preserving; thus, the *interlingua*, which is simply a semantic derivation tree (D-tree), has an isomorphic correspondence with the meaning-preserving operations in the syntactic derivation tree. Related to this point, the authors state that the interlingual/transfer debate is misplaced, and that it should be replaced by the issue of semantics-based (i.e., meaning-preserving mappings) versus form-based (i.e., structure-preserving mappings); they argue that all operations must necessarily preserve meaning regardless of the overall translation design.

Part II (chapters 6–9): *Elaboration of the Method*. In this part, the translation components are described in more detail. Rosetta includes a reversible morphology component, a set of dictionaries and associated partial acquisition techniques, and a syntactic component consisting of S-rules and M-rules (i.e., mappings from lexical trees to S-trees and from S-trees to D-trees, respectively). Chapter 9 introduces the notion of *modular and controlled M-grammars*, a key component of the solutions proposed for handling complex phenomena in subsequent chapters. Up to this point, it is assumed that the M-grammar rules apply freely and that no distinction is made between meaning-preserving rules (i.e., rules that preserve semantic content of main predicates such as “go”) and nonmeaningful rules (i.e., transformational rules that are semantically content-free such as “do-insertion” in yes/no questions). This chapter describes the negative consequences of these two assumptions and then presents an extension to the framework that allows certain difficulties to be avoided. The extended framework divides M-grammars into subgrammars, each corresponding to a different phrasal unit (e.g., NP) and each associated with a pre-specified control expression that defines an ordering of rule application.

Part III (chapters 10–11): *Linguistic Aspects*. This part of the book elaborates on the description given in chapter 9. The main point is that controlled M-grammars make it possible to find a proper balance between purely syntactic requirements and the requirements of a compositional grammar. Chapter 10 demonstrates that the combi-

² Despite the high level of technical detail, I found only two typographical errors in the book, both of them completely benign.

nation of meaning-preserving rules and syntactic transformations makes the relation between form and meaning indirect. Chapter 11 shows that the use of transformational rules (based on Chomsky 1957, 1965) makes it possible to incorporate and extend analyses of complex phenomena in a systematic and direct way.

Part IV (chapters 12–16): *Solution to Translation Problems*. This part, which arrives a bit late on the scene, presents what could be considered the main contribution of the Rosetta project. A wide range of translation problems, and their corresponding solutions, are discussed in great detail. Chapters 12 and 13 present an approach to the divergence problem that is contrasted with that of a number of MT researchers, in particular, work on categorial, promotional, and demotional divergences (Dorr 1993), structural mismatches (Estival, Ballim, Russell, and Warwick 1990), category changes and head switching (Lindop and Tsujii 1991; Kaplan and Wedekind 1993), embedding differences (Sadler and Thompson 1991; Kaplan, Netter, Wedekind, and Zaenen 1989), and complex lexical mismatches (Thurmair 1990). Chapters 14–16 present an approach to handling three additional classes of translation phenomena: temporal expressions, idioms and complex predicates, and scope and negation. These are given a much more comprehensive treatment than that of any other published work I have encountered in the MT field. A number of issues raised in part IV of the book deserve more detailed discussion; see section 2 below.

Part V (chapters 17–20): *Formal Aspects and Implementation*. The first two chapters in this part, while justifiably necessary, are a bit tedious to read. Two different perspectives on the definition of M-grammar components and their relation to the Rosetta modules are presented: a mathematical formalization and an attribute grammar view. These chapters are somewhat incongruent with the rest of the book when one considers the broader goals established at the outset—they serve only to demonstrate that: (a) properties such as strict isomorphism between M-grammars cannot be proved; and (b) the surface grammar of the syntactic component can be viewed as an attribute grammar, thus providing a framework for a cubic parsing algorithm. Chapter 19 presents a more interesting perspective than that of the first two chapters; it describes an algebraic view, not of the Rosetta system itself, but of its main ingredient, the *compositional* translation method. A formal description of the method's essential characteristics is provided; this is used to investigate the power of the method and also to compare the approach with that of other researchers. An interesting point that falls out of chapter 19 is that, because the transfer between the SL and TL is an isomorphism, the interlingua could be viewed as a nonessential aspect of the system; this point will be addressed further in section 2.

Some weaknesses are revealed in chapter 20 (Part V) and also in chapter 21 (the Conclusion). A brief description of software-engineering aspects and results is provided, yet there are almost no details about the implementation, i.e., how the different modules of Rosetta are integrated, which programming language is used, what the extent of user interaction is, etc.³ The authors simply state that such design decisions are “a matter of taste rather than of principle” (page 417). There is also no information about execution times for the numerous complex phenomena presented in the first 19 chapters. The omission of these details makes it difficult to evaluate the system, especially since so many of Rosetta's characteristics (e.g., the isomorphism requirement) depart radically from those of fully running and well-tested systems (e.g., the LOGOS system (Gdaniec 1994)). Although the developers describe an experiment where the approach is evaluated using a back-translation technique, it is not clear that this tech-

3 See Hutchins and Somers (1992, pages 279–296) for a more detailed discussion of this point.

nique can be used as a valid test of the system and, in any case, the reported results are not encouraging; these points are discussed further below.

2. Issues

We now direct our attention toward three issues:

- Status of interlinguality and adherence to the isomorphism principle.
- Handling of certain divergences within the compositional framework.
- Metrics used for evaluation of the system.

As mentioned above, one could view the interlingua (i.e., the semantic derivation tree) as a nonessential aspect of the system. In order to conform to the Principle of Isomorphism, the Rosetta designers are forced to impose the restriction that the semantic derivation tree contains no information corresponding to non-meaningful rules of syntax. However, the nodes of the semantic derivation tree are not assigned universal status; rather, these stand only as placeholders for meaning-preserving rules in the syntax module. Thus, the same effect could be achieved by a simple transfer mapping between analogous meaning-preserving rules in the SL and TL.

Indeed, this point has been brought up previously (Hutchins and Somers 1992, p. 294), where the primary criticism is that the SL and TL grammars are so tightly attuned that the approach might even be considered direct MT. Given that the translation operations in Rosetta induce complex transformations (e.g., for handling head-swapping divergences), I do not fully agree with Hutchins and Somers's assessment; however, I see no advantage to assuming an interlingual representation (at least one that does not have universal status), especially given that the same set of phenomena has been handled previously in simpler transfer designs.

Rosetta's separation of meaning-preserving and non-meaningful rules is reminiscent of the notion of relaxed compositionality in transfer systems such as Eurotra and its descendant, MiMo (Arnold, Krauwer, des Tombe, and Sadler 1988; Arnold and Sadler 1990), where "regular" rules are used for compositional phenomena and "exceptional" rules are used for noncompositional phenomena. (Not surprisingly, both groups address the same phenomena, e.g., head-swapping translations.) In fact, the authors suggest (pages 59–60) that the input-output behaviors of both designs are entirely equivalent. The only motivation for using an interlingua is that it is preferable to have an explicit stage where the meaning of the input expression is accessible for interactive disambiguation. This may be so, but because there is no discussion of an interactive disambiguation component, the full benefit of this design is not entirely obvious.

Regarding the second issue, that of handling divergences, it is clear that the Rosetta researchers have undertaken a thorough investigation in this area. However, some of their solutions rely heavily on the Principle of Compositionality, which, as we will see, has its limitations.

Consider the case of categorial divergence in which two phrasal heads are swapped:

English:	Mary happened to come
Dutch:	Mary kwam toevallig 'Mary came by chance'

Invariably, interlingua designers (myself included) are forced to select a canonical form that parallels one language or the other in the case of head-swapping phenomena.⁴ In the example above, the representation that is chosen corresponds logically to the canonical form ‘by-chance(Mary, come)’. Thus, the English syntactic structure parallels the canonical form: the verb construct corresponding to ‘by-chance’ (*happen to*) takes as its argument the clause corresponding to ‘(Mary, come)’ (*came*). The Dutch syntax, on the other hand, is not in synch with the canonical form since the verb corresponding to ‘(come, Mary)’ (*kwam*) takes as its argument the adverbial construct corresponding to ‘by-chance’ (*toevallig* or ‘by chance’).

In order to handle such cases compositionally, the subgrammar approach (see description in part II above) must be further revised so that, within the application of a subgrammar (e.g., ADVP), a “switch rule” is invoked and normal processing is interrupted; control is then passed to a different subgrammar (e.g., VP) and a new category (derived from an *argument* of the canonical head) takes over the role of syntactic head. Given that the book omits many implementational details, it is not clear how a “switch rule” is triggered. One assumption that is consistent with the framework described in chapter 9 is that interrupts are specified as part of the control expression associated with each subgrammar. This leaves open the question of how grammar-driven interrupts interact with idiosyncratic requirements of individual lexical items. In the case above, the nonhead constituent *toevallig* could be viewed as a “deviant” in that it takes on nonhead status in the syntax but head status in the canonical form. It would make a great deal of sense to encode such information in the lexicon (i.e., associate a lexical marker with the word itself as in Dorr (1993)). With such an encoding, it would not be the case that *every* Dutch adverbial (e.g., *gisteren* ‘yesterday’) triggers a grammar interrupt; only *certain* adverbials would act as triggers, namely those associated with a lexical marker (e.g., *toevallig*).

A more troubling aspect of the “switch-rule” approach is that it is difficult, or perhaps impossible, to accommodate head-swapping cases where the “deviant” serves as a head in the syntax but a nonhead in the canonical form. Consider the following example:

English: Mary usually goes to school
Spanish: Mary suele ir a la escuela
 ‘Mary is accustomed to go to school’

In this example, the canonical form could, arguably, be ‘go(Mary, school, usually)’.⁵ In this case, the English syntax parallels the canonical form. By contrast, the Spanish syntax includes the “deviant” verb *suele*, which is a head in the syntax but a nonhead in the canonical form; this is the inverse of the previous example. The head-swapping cases on pages 249–250 of the book do not include such a case, perhaps because this would force an interrupt to occur too late—*after* the syntactic structure corresponding to the logical head has already been built within a VP subgrammar. It seems that the choice of canonical form is rigid in that a potential “deviant” (such as *suele* in the current example and *toevallig* in the previous example) must always be associated

⁴ My preference would be to argue for the choice of interlingua on linguistic grounds (e.g., tests that predict lexical-semantic behavior such as those in Levin (1993), especially in a project like Rosetta, where the developers claim to be modeling linguistic knowledge. However, choosing an interlingua is undeniably a slippery endeavor (see, for example, Dorr and Voss (1993)) and so I am more than willing to leave this issue aside for the purposes of this review.

⁵ See Dorr (1993) for linguistic justification of this representation.

with the head in the canonical representation.⁶ Thus, even though it might not be linguistically justified, the canonical representation for the Spanish sentence would be ‘be-accustomed(Mary, go(school))’. By contrast, the English sentence would map into the canonical form given above, which means that the two would never be translation equivalents. Instead the system would force the following, more literal, translation pairs (in both directions):

English: Mary usually goes to school
Spanish: Mary usualmente va a la escuela

English: Mary is accustomed to going to school
Spanish: Mary suele ir a la escuela

The above difficulties are symptomatic of a more far-reaching problem—lexical-semantic information is ignored in the compositional derivation of the interlingua. Taking a grammar-driven approach forces the Rosetta developers to regard cases of lexical-semantic mismatch as purely grammatical. For example, in chapter 12, the authors view *conflation* as a simple one-to-many correspondence (e.g., the relation between the English word *love* and the equivalent multi-word phrase *amar a* in Spanish). This is not at all the conventional view as presented by Talmy (1983, 1985), where conflation is considered to be a much more complex notion, i.e., a meaningful word in one language (e.g., the English verb *float*) is composed of semantic units that correspond to more than one meaningful word in the other language (e.g., *mudaba flotando* or ‘move floating’). This version of conflation is not adequately addressed in the book, most likely because the compositional framework does not extend to the lexical level. In fact, it is not clear whether such an extension would be possible in the context of the isomorphism restriction, which impedes the mapping of a single meaningful constituent in one language to more than one meaningful constituent in another language. The only way to finesse this would be to render one constituent “meaningless.” But then which one would we pick? Suppose, for example, that we pick *mudaba* to be the meaningless constituent in the phrase *mudaba flotando*—then how would we prevent the verb *mudar* from being ignored in contexts where it stands alone, as in *John se mudó a las afueras de Boston* (or ‘John moved outside of Boston’)? The possibility of extending the notion of compositionality into the lexicon is an issue that should be discussed in a book on a compositional MT system, especially given the wealth of literature in this area (see, for example, Levin (1993; Levin and Rappaport Hovav 1995)).

Finally, we turn to the issue of the metrics used for evaluation of the Rosetta system. The authors report on an experiment in which 1246 Dutch sentences are first translated into English and then back into Dutch; the Dutch input and output are then compared, with a result of 57% correctness. The reliability of the back-translation technique is questionable, at best. Given that the development of evaluation metrics for MT has been an area of serious debate for at least three decades, it is surprising that the Rosetta researchers so readily accept this as a true test of the system.⁷

The reason I am skeptical about this technique is that it provides very little insight into the quality of the translation system. If the Dutch input and output are the same,

⁶ On page 249, the authors mention that the label “categorical divergence” covers both the promotional and demotional cases in Dorr (1993); no explanation is given for this move, but I would suspect it is because of this rigidity.

⁷ See, for example, the 1993 special issue of *Machine Translation* on evaluation of MT systems; in particular, the article by Arnold, Sadler, and Humphreys (1993) provides a thorough overview of evaluation techniques, none of which includes back-translation.

we know nothing about the quality of the intervening English sentence. We simply know that the system has faithfully maintained the isomorphism requirement. If the Dutch input and output are different, we still know nothing about the quality of the intervening English sentence because the “best” translation of that sentence is not necessarily the Dutch original. There might even be cases where the back-translated sentence is an improvement on the original input. To illustrate this point consider the following case where an English sentence is back-translated via Spanish:

English: Mary is accustomed to going to school
Spanish: Mary suele ir a la escuela
English: Mary usually goes to school

Here the English input does not match the final English output, yet the translation is clearly valid at each step.

Because of the isomorphism restriction in Rosetta, it is doubtful that the low rate of 57% can be attributed to cases that parallel the example given here. In fact, it is surprising that Rosetta does so poorly despite the isomorphism requirement. One would expect results to be spuriously high given that there are surely cases where the Dutch input and output match each other, yet the intervening English sentence—albeit structurally isomorphic to the Dutch—is not an appropriate translation. Other than a brief mention of certain phenomena that were not yet handled (e.g., coordination), the authors do not speculate as to why the result rate is so low or, even worse, whether this rate is accidentally high due to the isomorphism requirement.

3. Summary

The book is of great interest to the computational linguistics and machine translation communities and could, perhaps, serve as a textbook for instructors in these areas. Its weaknesses are that it omits discussion of the role of the lexicon in the handling of certain divergence types and that it fails to provide a thorough evaluation of the techniques that are proposed. Even so, the authors have written clearly about some of the hardest phenomena in MT and have remained consistent in their handling of these phenomena under strict assumptions of compositionality and isomorphism; their treatment is more in-depth and broad-scope than any other MT book I have encountered.

Acknowledgments

Thanks to Clare Voss for her useful comments and insights and to Bruce Dawson and Doug Jones for reading and suggesting revisions on an earlier draft of this review.

References

Appelo, Lisette (1993). *Categorial divergences in a compositional translation system*. Doctoral dissertation, University of Utrecht, Utrecht.

Arnold, Doug; Krauwer, Steven; des Tombe, Louis; and Sadler, Louisa (1988). “Relaxed compositionality in machine translation”. *Proceedings, Second International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages*, Carnegie Mellon University, Pittsburgh, PA.

Arnold, Doug, and Sadler, Louisa (1990). “Theoretical basis of MiMo.” *Machine Translation*, 5(3), 195–222.

Arnold, Doug; Sadler, Louisa; and Humphreys, R. Lee (1993). “Evaluation:

- An assessment." *Machine Translation*, 8(1-2), 1-24.
- Chomsky, Noam A. (1957). *Syntactic structures*. Mouton: The Hague.
- Chomsky, Noam A. (1965). *Aspects of the theory of syntax*. Cambridge, MA: The MIT Press.
- Dorr, Bonnie J. (1993). *Machine translation: A view from the lexicon*. Cambridge, MA: The MIT Press.
- Dorr, Bonnie J., and Voss, Clare R. (1993). "Machine translation of spatial expressions: Defining the relation between an interlingua and a knowledge representation system." *Proceedings, Twelfth Conference of the American Association for Artificial Intelligence*, Washington, D.C., 374-379.
- Estival, Dominique; Ballim, Afzal; Russell, Graham; and Warwick, Susan (1990). "A syntax and semantics for feature-structure transfer." *Proceedings, Third International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages*, Linguistics Research Center, The University of Texas, Austin, TX.
- Gdaniec, Claudia (1994). "The LOGOS translatability index." *Proceedings, First Annual Association for MT in the Americas Conference on Partnerships in Translation Technology*, Columbia, MD, 97-105.
- Hutchins, W. John, and Somers, Harold (1992). *An Introduction to Machine Translation*. London, England: Academic Press.
- Janssen, Theo (1986). *Foundations and Applications of Montague Grammar: Part I, Philosophy, Framework, Computer Science*. Doctoral dissertation, Mathematical Centre, Amsterdam.
- Kaplan, Ronald M.; Netter, Klaus; Wedekind, Jürgen; and Zaenen, Annie. (1989). "Translation by structural correspondences." *Proceedings, Fourth Conference of the European Chapter of the Association for Computational Linguistics*, Manchester, 272-281.
- Kaplan, Ronald M., and Wedekind, Jürgen (1993). "Restriction and correspondence-based translations." *Proceedings, Sixth Conference of the European Chapter of the Association for Computational Linguistics*, Utrecht.
- Levin, Beth (1993). *English verb classes and alternations: A preliminary investigation*. Chicago, IL: University of Chicago Press.
- Levin, Beth, and Rappaport Hovav, Malka (1995). *Unaccusativity: At the syntax-semantics interface*. Cambridge, MA: The MIT Press.
- Lindop, Jeremy, and Tsujii, Jun-ichi (1991). "Complex transfer in MT: A survey of examples." Technical Report CCL/UMIST Report 91/5, Center for Computational Linguistics, UMIST, Manchester.
- Sadler, Louisa, and Thompson, Henry S. (1991). "Structural non-correspondence in translation." *Proceedings, Fifth Conference of the European Chapter of the Association for Computational Linguistics*, Berlin, 293-298.
- Talmy, Leonard (1983). "How language structures space." In *Spatial orientation: Theory, research, and application*, edited by H.L. Pick and L.P. Acredolo, 225-282. New York: Plenum Press.
- Talmy, Leonard (1985). "Lexicalization patterns: Semantic structure in lexical forms." In *Language typology and syntactic description, volume 3: Grammatical categories and the lexicon*, edited by T. Shopen, 57-149. Cambridge, England: Cambridge University Press.
- Thurmair, Gregor (1990). "Complex lexical transfer in METAL." *Proceedings, Third International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages*, Linguistics Research Center, The University of Texas, Austin, TX. 91-107.

Bonnie J. Dorr received her Ph.D. in computer science from the Massachusetts Institute of Technology in 1990. She is co-director of the University of Maryland Computational Linguistics and Information Processing (CLIP) Laboratory and has recently received the Alfred P. Sloan Research Fellowship Award and an NSF Young Investigator Award for her work on large-scale interlingual machine translation. Dorr's address is: Department of Computer Science, A.V. Williams Building, University of Maryland, College Park, MD 20742; e-mail: bonnie@cs.umd.edu.