Mach Translat (2005) 19:213–227
DOI 10.1007/s10590-006-9009-3

ORIGINAL PAPER

MT model space: statistical versus compositional versus example-based machine translation

Dekai Wu

Received: 23 January 2006 / Accepted: 25 August 2006 /
Published online: 14 February 2007

Abstract We offer a perspective on EBMT from a statistical MT standpoint, by developing a three-dimensional MT model space based on three pairs of definitions: (1) logical versus statistical MT, (2) schema-based versus example-based MT, and (3) lexical versus compositional MT. Within this space we consider the interplay of three key ideas in the evolution of transfer, example-based, and statistical approaches to MT. We depict how all translation models face these issues in one way or another, regardless of the school of thought, and suggest where the real questions for the future may lie.

Keywords Statistical MT · Example-based MT · Compositional MT

1 Introduction

We offer a perspective on example-based machine translation (EBMT) vis-à-vis statistical machine translation (SMT), and traditional compositional rule-based machine translation (RBMT). Our discussion is partly motivated by the fact that in common usage we find two very different kinds of senses to terms such as "example-based MT", "statistical MT", and the like. Some have argued for fuzzy *sociocultural* senses, which arise and mutate in accord with the historical evolution of research communities and subcommunities. In contrast, we believe it is important to adhere to the *formal* senses, which are well-defined technical concepts that are mathematically precise and can be objectively tested.

Where this distinction becomes important is in the context of the following sorts of frequently asked questions: What is the definition of EBMT? Do we even know what EBMT is? Is there a strict definition of EBMT, or are there simply a large number

D. Wu (✉)
Department of Computer Science, HKUST Human Language Technology Center, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong
e-mail: dekai@cs.ust.hk

of different models all using corpora, rules, and statistics to varying degrees? Is X a kind of EBMT model? Does X's model qualify as EBMT but not SMT? Are all SMT models (perhaps excluding the IBM models) actually EBMT models as well? Are EBMT models actually SMT models? Can a rule-based model be example-based? Statistical? (Often, one senses a loudly implied "just" in such questions.)

Formal definitions, rather than sociocultural ones, are necessary when it comes to questions like these. Otherwise the questions tend toward meaningless quibbles: asking whether X is a kind of EBMT model easily degenerates into an exercise in sociocultural analysis. Thanks to constant advances, shifts, cross-pollination, and hybridization in the modeling approaches of any active subcommunity, anything new a subcommunity embraces rapidly becomes part of its definition. Thus, the sociopolitical senses are always vaguely defined moving targets.

On the other hand, by appealing to formal definitions when answering such questions, we enable rigorous advances in our understanding of the underlying mathematical regularities and properties across various MT models, regardless of which subcommunity they may have arisen in. This serves to eliminate superficial differences that often impede progress due to antagonisms of vocabulary.

Accordingly, as a foundation for our discussion, we first consider in the following sections how to define formally SMT, EBMT, and compositional MT. The definitions we arrive at suggest a three-dimensional "MT model space" whose axes correspond to the three formal definitions. Given this MT model space, we then consider the trajectory of historical development and interplay of ideas of the three fuzzy sociocultural strands often loosely referred to as SMT, EBMT, and RBMT research subcommunities.

2 What is SMT?

To understand what "statistical MT" means, one begins by asking what "statistical" means. Given that the term is long established, one may naturally consult dictionaries, which all yield pretty much the same answer:

statistical *adj*

1. of, relating to, based on, or employing the principles of statistics (Merriam-Webster 2002);
2. of or relating to statistics; "statistical population" (WordNet 2003);
3. of, relating to, or employing statistics or the principles of statistics (American Heritage 2000);
4. relating to statistics (OED 2005).

So what, then, is the definition of "statistics"?

statistics *n*

1. a branch of mathematics dealing with the collection, analysis, interpretation, and presentation of masses of numerical data (Merriam-Webster 2002);
2. a branch of applied mathematics concerned with the collection and interpretation of quantitative data and the use of probability theory to estimate population parameters (WordNet 2003);

3. the mathematics of the collection, organization, and interpretation of numerical data, especially the analysis of population characteristics by inference from sampling (American Heritage 2000);
4. numerical data (American Heritage 2000);
5. science of collecting and classifying a group of facts according to their relative number and determining certain values that represent characteristics of the group (Columbia 2005);
6. the collection and analysis of numerical data in large quantities (OED 2005).

We may reasonably conclude:

statistical MT *n* MT making nontrivial use of mathematical statistics and probability. *syn* probabilistic MT. *ant* (purely) logical MT
logical MT *n* MT making nontrivial use of mathematical logic (without statistics and probability). *ant* statistical MT

Note that since probability theory is founded on top of set theory and propositional logic, all SMT models are inherently logical MT models, but in addition they are augmented with a measure that assigns probabilities to the sets. In other words, both statistical and logical models are symbolic, depending in their vocabulary and structure on symbols and variables. (The converse does not hold, however; logical MT models are not inherently SMT models.)

Just as there are many possible approaches to MT making use of the various models in mathematical logic (using propositional vs. first-order vs. modal vs. nonmonotonic calculi, for segmentation vs. parsing vs. disambiguation components, in direct vs. transfer vs. interlingual vs. knowledge-based architectures, with deductive vs. abductive vs. inductive engines, and so on), so too are there many possible approaches to MT making use of the various models in mathematical statistics, and to widely varying degrees of sophistication. In addition to the basic methods of descriptive statistics, various *statistical inference* methods of predictive statistics are of particular relevance to MT and other prediction tasks where the objective is optimal decision making. Even here, it is noteworthy that many alternative approaches exist (using generative vs. discriminative vs. minimum Bayes risk formulations, for segmentation vs. parsing vs. disambiguation components, in direct vs. transfer vs. interlingual vs. knowledge-based architectures, under classic vs. Bayesian vs. nonparametric assumptions, and so on). Speaking generally, statistical inference is of course almost always ultimately used for disambiguation, but this can take place in either decision or search models at various levels, and can be used for ranking, scoring, parameter estimation, etc. Occasionally one sees attempts to pigeonhole SMT as one particular model or some narrow class of models, say, the IBM models (Brown et al. 1990) or source-channel models, or Bayesian decision models. However, to draw an analogy to models based on symbolic logic, this would be akin to the obvious mistake of trying to pigeonhole "logical MT" as encompassing only one particular rule-based architecture or one particular logic.

3 What is EBMT?

Although it has not often been explicitly mentioned in the EBMT literature since Nagao (1984) first proposed "translation by analogy" (cf. Lepage and Denoual 2005), example-based methods belong to the tradition of analogical models arising in the mid-1980s under various similar names including “case-based reasoning” (CBR) as in Kolodner (1983a,b), “exemplar-based reasoning” as in Porter and Bareiss (1986) or Kibler and Aha (1997), “instance-based reasoning” as in Aha et al. (1991), “memory-based reasoning” as in Stanfill and Waltz (1988), or “analogy-based reasoning” as in Hall (1989) or Veloso and Carbonell (1993). One detailed point-by-point elaboration of EBMT as an application of CBR is given by Collins and Somers (2003). Analogical reasoning at translation runtime is also identified by Turcato and Popowich (2003) as the main distinguishing property of EBMT approaches. Broad EBMT surveys showing the centrality of reasoning by analogy to examples are found in Somers (1999, 2003), and Hutchins (2005a) (cf. also Hutchins 2005b).

The distinction between EBMT and other corpus-based learning models is, thus, essentially an instance of the distinction between CBR and other machine learning models. As with all the analogical methods, the term “example-based” does not encompass all corpus-based, data-driven, or learning methods. Rather, it has a more specific implication about how and when learning and adaptation take place, as follows.

The key defining characteristic distinguishing these models from others is that they make nontrivial use of a large library of examples/cases/exemplars/instances at runtime, that is, during the task performance/testing phase rather than the learning/training phase. New problems are solved at runtime via analogy to similar examples retrieved from the library, which are broken down, adapted, and recombined as needed to form a solution. This stands in contrast to most other machine learning approaches which focus on heavy offline learning/training phases, so as to compile or generalize large example sets into abstracted performance models consisting of various forms of abstracted schemata (which are normally much smaller than the entire set of training examples).

Leaning toward memorization rather than abstraction of the training set makes some significant tradeoffs. On one hand, given sufficiently large example libraries, memorization avoids loss of coverage often caused by incorrect generalization or overgeneralization. In the extreme case, memorization approaches are guaranteed to reproduce exactly all unique sentence translations from the training corpus, something abstracted schematic approaches may not necessarily do. On the other hand, memorization approaches tend to undergeneralize, and runtime space and time complexity are vastly increased.

Note that the ramifications of an example-based approach primarily manifest themselves only in a resource-bounded view (Russell and Wefald 1991). If memory space and run time are unbounded, then in theory it would make no difference whether generalization and/or adaptation occur during training vs. testing. In practice, of course, it makes a great deal of difference.

We see the following definitions, then:

example-based MT *n* MT making nontrivial use of a large library of examples during translation runtime (i.e., testing as opposed to training). *syn* case-based MT. *ant* schema-based MT
schema-based MT *n* MT making nontrivial use of abstract schemata during translation runtime (without a large library of examples). *ant* example-based MT

In practice, entirely “pure” systems do not exist. All systems actually make some form of tradeoff in how much offline preprocessing they do during learning/training phases, vs. how much online processing they defer until translation runtime/testing. Likewise, various granularities of schemata are possible. Thus, as we discuss below, "example-based" and "schema-based" anchor the endpoints of a continuum rather than a simple binary classification.

4 What is compositional MT?

Both SMT and EBMT are data-driven methods that stand in contrast to traditional approaches of building MT models out of compositional rules (traditionally by hand as in Systran (Toma 1976), though this is not strictly necessary). The oft-heard term "rule-based MT" is rather misleading in the context of seeking formal definitions since, mathematically speaking, all MT models use some form of rules, even probabilistic and/or example-based models. It is the *compositionality* of rules that is really the intended emphasis. Leaving aside for now considerations of semantics and pragmatics, compositional MT models employ bilingual "transfer rules" or "translation rules" or "transduction rules" that declaratively describe how larger chunks can be translated by recursively composing smaller translated chunks.

Note that transduction/translation/transfer rules can also of course be used to hold translations of purely lexical collocations or phrases, as a special case. This is especially common for the "reordering rules" in direct translation models such as Météo (Chandioux 1976) or Systran (Toma 1977). However, the distinction between direct and transfer MT is extremely hazy; as Hutchins and Somers (1992, p75) observe, "strictly speaking all translation systems involve 'transfer' of some kind, the conversion of a source text or representation into a target text or representation." Whether a system is considered direct or transfer MT essentially depends on the fuzzy criterion of how much or how little language-specific monolingual analysis is performed, that is, how closely the intermediate representations resemble the source and target texts themselves.

We add the following definitions:

compositional MT *n* MT making nontrivial use of compositional transfer/ transduction rules. *syn* transfer MT. *ant* lexical MT
lexical MT *n* MT making nontrivial use of lexical transfer/transduction rules (without compositional rules). *ant* compositional MT

5 MT model space

What is interesting to note in the foregoing discussion is that there is nothing mutually exclusive between the definitions of SMT, EBMT, or compositional MT. Rather, they focus on tackling independent issues that *any* MT model faces.

This suggests that we may view any MT model as sitting at some point within a three-dimensional space defined by axes corresponding to the degree of statistical, example-based, and compositional techniques employed, as depicted in Fig. 1.

The example-based x-axis represents the degree to which abstraction (generalization and/or adaptation) is performed during testing, as opposed to during training. Models vary along the spectrum from schema-based models which emphasize abstraction of the training set, to example-based models which emphasize memorization of the training set.

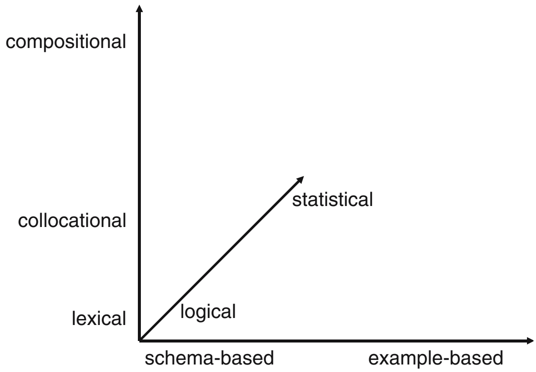

Fig. 1 The space of MT models (see Sect 5)

The compositional y-axis represents the degree to which rules are compositional, as opposed to lexical. Models vary along the spectrum from flat lexical models, to fully recursive compositional models. Collocational/phrasal models fall somewhere in the middle, since they emphasize composition of lexical items, but often without the full use of categories.

The statistical z-axis represents the degree to which models make appropriate use of statistics, as opposed to logic and set-theoretic models. Models vary along the spectrum from purely logical models, to models that make increasing use of statistics and statistical inference.

Two extreme points in this space are highlighted in Fig. 2. At the origin lie traditional word-to-word "translation lexicons," which are logical, lexical, and schema-based. In fact, any decent real translation lexicon contains numerous collocational or phrasal translations such as *Hong Kong*/香港, and therefore lies somewhere between lexical and collocational.

Similarly, translation memories such as TRADOS or Déjà Vu are logical, as well as typically somewhere between lexical and collocational, but they are example-based rather than schema-based. A translation memory stores many examples of past lexeme or collocation translations, unlike a translation lexicon where each lemma constitutes an abstract schema.

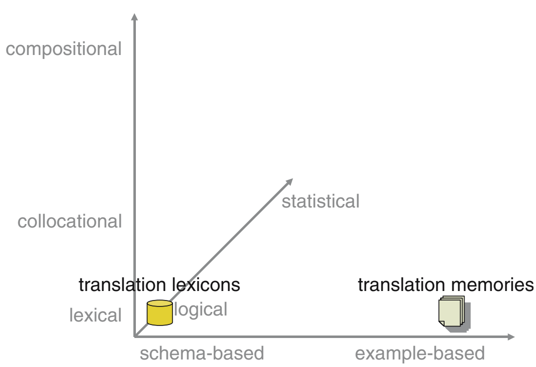

Fig. 2 Two anchors in the space of MT models: translation lexicons and translation memories

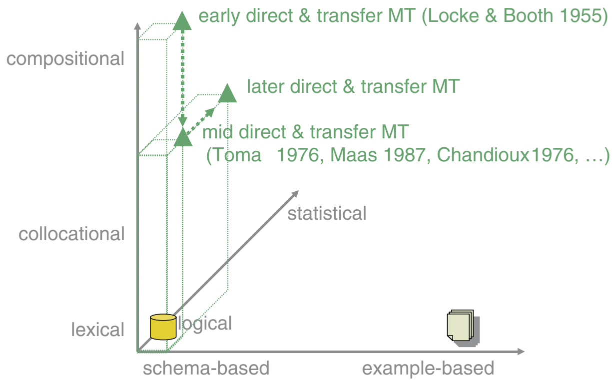

Fig. 3 Historical trajectory of development of direct and transfer MT models

Let us consider the historical development of direct and transfer MT models, viewed as a trajectory in the three-dimensional space as depicted in Fig. 3. In the early MT era of the Georgetown-IBM demonstration, it was thought that grammar would play the central role, so the emphasis was focused on abstract compositional rules operating over word-to-word translation (Locke and Booth 1955). Statistics did not yet play a great role (although cryptographic methods and word frequency studies were envisioned ever since Weaver 1949), and computational resources were too limited to permit large libraries of examples at runtime. However, it quickly became clear that simplistic word-to-word translation could not translate even elementary collocations like *Hong Kong*/香港, so collocations were added to form phrasal translation lexicons; even early Systran (Toma 1976), SUSY (Maas 1987), Météo (Chandioux 1976) and GETA's Ariane (Vauquois 1975, Vauquois and Boitet 1985) architectures incorporated idiom/compound/collocation dictionaries. Subsequent development included simple numerical scoring functions in preference systems to disambiguate on the basis of frequency, as for example in SUSY (Maas 1987). Today nearly all direct and transfer MT models are adopting increasing use of at least simple statistical methods, including for example Systran (Senellart et al. 2003).

We now turn to the trajectory of historical development of EBMT models. As depicted in Fig. 4, the early EBMT models following Nagao (1984) focused on translation by analogy against a large library of translation examples for lexical collocations as in Sumita and Iida (1991) and updated in Sumita (2003), with fairly ad hoc numerical measures. Modern versions of this "pure" EBMT approach include Andriamanankasina et al. (2003), Bond and Shirai (2003), and Lepage and Denoual (2005). Subsequent development pushed in different directions. Lexical collocation translations were augmented by abstracted templates containing variables (i.e., transduction rules) thereby simultaneously moving in two dimensions toward both compositional and schema-based approaches, as in Kitano and Higuchi (1991), Furuse and Iida (1992), Kaji et al. (1992), or Matsumoto et al. (1993), and subsequently furthered in work such as Cicekli and Güvenir (1996, 2003), Veale and Way (1997), Brown (1999, 2003), McTait and Trujillo (1999), Carl (2003), Yamamoto and Matsumoto (2003), Groves et al. (2004), Way and Gough (2005), Cicekli (2005), and Groves and Way (2005b). At the same time, gradually increasing use of probabilities in similarity metrics and in scoring adaptation and composition of hypotheses has also moved EBMT

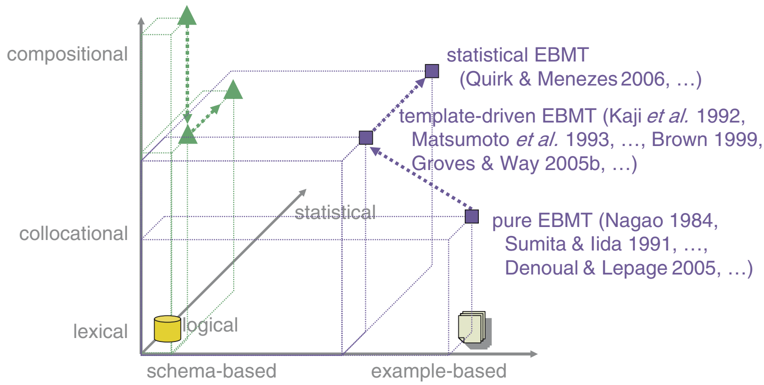

Fig. 4 Historical trajectory of development of EBMT models

in the direction of statistical approaches. Modern EBMT systems incorporate both; for example, Aramaki et al. (2005), Langlais and Gotti (2006), Liu et al. (2006), and Quirk and Menezes (2006) aim for probabilistic formulations of EBMT in terms of statistical inference.

Where does the development of SMT fit in the picture? As depicted in Fig. 5, the first IBM source-channel models (Brown et al. 1990) began with very simple word-to-word lexical translation models: they were statistical, lexical, and schema-based. The later IBM models attempted to model lexical collocation translation effects using indirect means (Brown et al. 1993).

With the introduction of tree-structured SMT approaches supporting stochastic transduction rules, both lexical collocations and compositional structure could be explicitly incorporated into SMT, as in the bracketing inversion transduction grammar (BITG) models of Wu (1995a, 1996, 1997, 2000), and closely related subsequent models like those of Lu et al. (2001, 2002), Simard and Langlais (2003), Zens and Ney (2003), Zhao and Vogel (2003), Zens et al. (2004), Chiang (2005), Vilar and Vidal (2005), Wu (2005), and Wu and Fung (2005).

Tree-structured SMT models went on to incorporate more sophisticated compositionality, in the form of transduction rules that made use of abstract linguistic

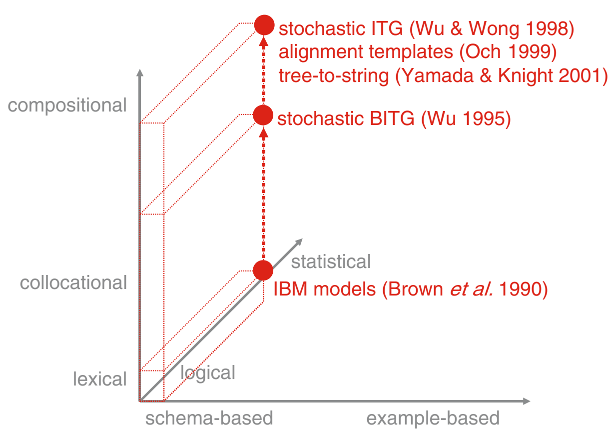

Fig. 5 Historical trajectory of development of SMT models (part 1)

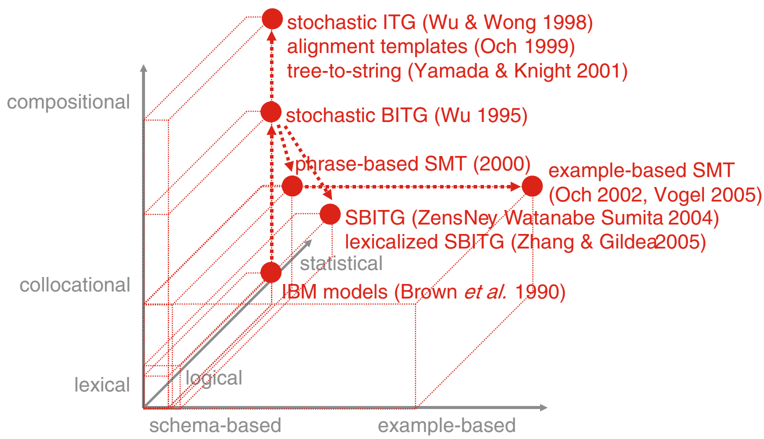

Fig. 6 Historical trajectory of development of SMT models (part 2)

categories, as in Wu (1995b), Alshawi et al. (1997, 1998, 2000), Wu and Wong (1998), Och et al. (1999), Yamada and Knight (2001), or Lin and Cherry (2003).

The subsequent development of tree-structured SMT witnessed two moves toward increased lexicalization: one focusing on purely lexical collocations, the other on lexicalizing the compositional SMT models. This is depicted in Fig. 6.

The first, occurring around the early 2000s, led to what one might think of as "example-based SMT". This move saw the incorporation of vastly expanded runtime libraries of well over 10^8 lexical collocation translation rules, essentially memorized from the training corpus, such as the phrase-based SMT models of Och and Ney (2002), Koehn et al. (2003), or Zhang and Vogel (2005).[1]

At the same time, expanded lexicons were also incorporated into recursively compositional SMT models, such as the stochastic BITG models of Zens et al. (2004) or the lexicalized stochastic BITG model of Zhang and Gildea (2005). The lexicons here, however, still do not memorize the training corpus to the same extent as EBMT models do.

6 Discussion

Carl (2000) proposed a "model of competence for corpus-based MT" that also classifies MT models along three dichotomies. Carl (2005) posits some similarities with our three-dimensional MT model space:

> The dichotomy *fine-grained* vs. *coarse-grained* coincides roughly with "schema-based" vs. "example-based" in Wu's MT space: translation units are likely to be coarse-grained when using a *large library of examples* ([cites this paper]) while they will be finer grained the more the schemas are abstracted. The axis *molecular* vs. *holistic* is related to "logical" vs. "statistical" since statistics can be

[1] Coming from the other direction, on a smaller scale, the hybrid example-based SMT model of Groves and Way (2005a, 2005b) adds about 430 k lexical collocation translation rules acquired by their EBMT system to a phrase-based SMT system that itself acquires roughly 1.73 m lexical collocation translation rules.

used to derive shades of meaning distinctions from corpora while mathematical logics are required to compose larger meaning entities from finite sets of features. The dimension *austere* vs. *rich* relates to "lexical" vs. "compositional" insofar as mere lexical translations will be close to their graphemic surface forms while rich representations are required for compositional translations. (Carl 2005)

However, upon closer inspection, we can see that the formal criteria in the definitions of Carl's dichotomies differ in important respects.

The first two of Carl's dichotomies are drawn from a theory of meaning proposed by Dummett (1975). First, a *molecular* theory of meaning "derives the understanding of an expression from a finite number of axioms" (Carl 2000, p997), whereas a *holistic* theory "derives the understanding of an expression through its distinction from all other expressions in that language" (*idem.*). Applied specifically to corpus-based MT, "in a molecular approach the meaning descriptions are obtained from a finite set of predefined features", whereas "in a holistic implementation meaning descriptions are derived from reference translations" (*ibid.*:1001).

At first blush, the formal criteria defining the molecular vs. holistic dichotomy might seem similar to the formal criteria defining the logical vs. statistical distinction. However, it is possible to build molecular models that either make use of mathematical statistics or make use only of mathematical logic, insofar as either kind of model can be designed such that the meaning descriptions are obtained from a finite set of axioms and/or predefined features. Similarly, holistic models need not be statistical. It is possible to conceive of a purely logical model that holistically derives meaning descriptions from an entire set of reference translations, say, in the vein of inductive logic models that rely solely on logical constraints without statistics or probability, such as the grammar induction model of Angluin (1980).

The second dichotomy is also from Dummett. An *austere* theory of meaning "merely relies upon simple recognition of the shape of the concepts", whereas "in a *rich* theory of meaning, the knowledge of the concepts is achieved by knowing the features of these concepts" (Carl 2000, p997). Carl applies this specifically to corpus-based MT saying "with an austere theory the system relies on the mere graphemic surface form of the text", whereas "in a system that uses a rich theory of meaning, complex representations are computed including morphological, syntactical, and semantical representations" (*ibid.*:1001).

Here too the formal criteria defining austere vs. rich diverge from those defining lexical vs. compositional. It is reasonably common, though not required, that compositional models tend to use richer representations. But the key differentiating characteristic that defines compositional models is that they permit recursive composition of structures, which is independent of whether an austere or rich model introspectively knows the features of these concepts.

The third dichotomy of Carl (2000) proposes that "a *fine-grained* theory of meaning derives concepts from single morphemes or separable words of the language, whereas in a *coarse-grained* theory of meaning, concepts are obtained from morpheme clusters" (*ibid.*: 997f). In terms of corpus-based MT, "in a fine-grained theory, the minimal length of a translation unit is equivalent to a morpheme while in a coarse-grained theory this amounts to morpheme cluster, a phrase or a sentence" (*ibid.*:1001).

The formal criteria defining fine-grained vs. coarse-grained are orthogonal to those defining schema-based vs. example-based. Whether the lengths of translation units are fine-grained morphemes or coarse-grained collocations/compounds/

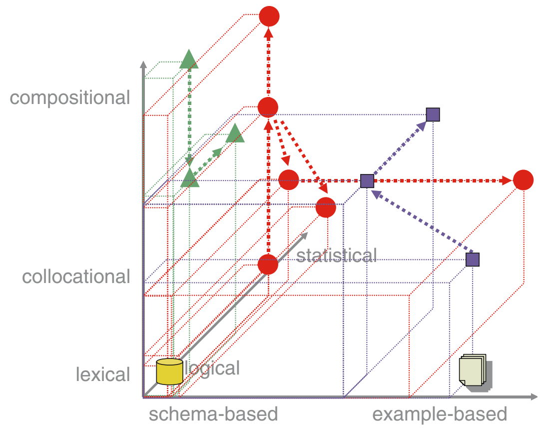

Fig. 7 Historical trajectories of development of transfer MT, EBMT, and SMT models

phrases/idioms is a distinction that lies on the compositional axis between lexical and collocational (since we consider collocations to be a one-level instance of composition). The units of translation may be either fine-grained morphemes or coarse-grained collocations, regardless of whether an example-based analogical model retrieves and adapts examples from a large library at translation runtime, or a schema-based model learns abstracted schemata during training and discards the concrete example base prior to seeing any test instances at translation runtime.

Thus Carl's dichotomies represent additional classification criteria for a typology of MT models, largely independent of the three dimensions we have discussed.

7 Conclusion

We have considered the interplay of three key ideas in the evolution of statistical, example-based, and transfer approaches to MT. The definitions of SMT, EBMT, and compositional MT we arrived at show that the issues actually constitute three independent dimensions, which all translation models face in one way or another. The historical development of the various schools of thought can thus be viewed as trajectories within a three-dimensional MT model space, as summarized in Fig. 7.

Numerous key research questions are highlighted by this visualization. While there is no clear division between the schools of thought, what is the right combination of ingredients? To what extent can compositional SMT models circumvent the need to memorize purely lexical collocations? Conversely, to what extent can large example libraries improve compositional SMT models? How much generalization from training examples is desirable? Independent of whether generalization is performed at training or testing time, what sort of generalization biases (similarity metrics, example selection criteria, adaptation and combination functions) perform best? Answers to these questions will ultimately help determine the optimum point in the MT model space for any given translation application.

Acknowledgements This work was supported in part by DARPA GALE contract HR0011-06-C-0023, and by the Hong Kong Research Grants Council (RGC) research grants RGC6083/99E, RGC6256/00E, and DAG03/04.EG09.

References

Aha DW, Kibler D, Albert MK (1991) Instance-based learning algorithms. Mach Learn 6:37–66
Alshawi H, Bangalore S, Douglas S (1998) Automatic acquisition of hierarchical transduction models for machine translation. In: COLING-ACL '98: 36th annual meeting of the Association for Computational Linguistics and 17th international conference on computational linguistics, Montreal, Quebec, Canada, pp 41–47
Alshawi H, Buchsbaum AL, Xia F (1997) A comparison of head transducers and transfer for a limited domain translation application. In: 35th Annual meeting of the Association for Computational Linguistics, Madrid, Spain, pp 360–365
Alshawi H, Douglas S, Bangalore S (2000) Learning dependency translation models as collections of finite-state head transducers. Comput Ling 26:45–60
American Heritage (2000) The American heritage dictionary of the English language, 4th edn. Houghton Mifflin, Boston, MA
Andriamanankasina T, Araki K, Tochinai K (2003) EBMT of POS-tagged sentences by recursive division via inductive learning. In: Carl and Way (2003), pp 225–252
Angluin D (1980) Inductive inference of formal languages from positive data. Inform Control 45:117–135
Aramaki E, Kurohashi S, Kashioka H, Kato N (2005) Probabilistic model for example-based machine translation. In: MT Summit X: The tenth machine translation summit, Phuket, Thailand, pp 219–226
Bond F, Shirai S (2003) A hybrid rule and example-based method for machine translation. In: Carl and Way (2003), pp 211–224
Brown PE, Cocke J, Della Pietra SA, Della Pietra VJ, Jelinek F, Lafferty JD, Mercer RL, Roossin PS (1990) A statistical approach to machine translation. Comput Ling 16:79–85; repr. in Nirenburg et al. (2003), pp 355–362
Brown PE, Della Pietra VJ, Della Pietra SA, Mercer RL (1993) The mathematics of statistical machine translation: Parameter estimation. Comput Ling 19:263–311
Brown RD (1999) Adding linguistic knowledge to a lexical example-based translation system. In: Proceedings of the eighth international conference on theoretical and methodological issues in machine translation (TMI-99), Chester, England, pp 22–32
Brown RD (2003) Clustered transfer rule induction for example-based translation. In: Carl and Way (2003), pp 287–305
Carl M (2000) A model of competence for corpus-based machine translation. In: Proceedings of the 18th international conference on computational linguistics: COLING 2000 in Europe, Saarbrücken, Germany, pp 997–1001
Carl M (2003) Inducing translation grammars from bracketed alignments. In: Carl and Way (2003), pp 339–361
Carl M (2005) A system-theoretical view of EBMT. Mach Translat 19:229–249
Carl M, Way A (eds) (2003) Recent advances in example-based machine translation. Kluwer Academic Publishers, Dordrecht, The Netherlands
Chandioux J (1976) Météo: Un système opérationnel pour la traduction automatique des bulletins météorologiques destinés au grand public [Météo: A working system for the machine translation of weather reports aimed at the general public]. META 21:127–133
Chiang D (2005) A hierarchical phrase-based model for statistical machine translation. In: 43rd annual meeting of the Association for Computational Linguistics, Ann Arbor, MI, pp 263–270
Cicekli I (2005) Inducing translation templates with type constraints. Mach Translat 19:283–299
Cicekli I, Güvenir HA (1996) Learning translation rules from a bilingual corpus. In: Second international conference on new methods in language processing (NeMLaP-2), Ankara, Turkey, pp 90–97
Cicekli I, Güvenir HA (2003) Learning translation templates from bilingual translation examples. In: Carl and Way (2003), pp 255–286
Collins B, Somers H (2003) EBMT seen as case-based reasoning. In: Carl and Way (2003), pp 115–153
Columbia (2005) The Columbia electronic encyclopedia. Columbia University Press, New York, NY

Dummett MAE (1975) What is a theory of meaning? In: Guttenplan S (ed) Mind and language. Oxford University Press, Oxford, England, pp 97–138
Furuse O, Iida H (1992) Cooperation between transfer and analysis in example-based framework. In: Proceedings of the fifteenth [sic] international conference on computational linguistics, COLING-92, Nantes, France, pp 645–651
Groves D, Hearne M, Way A (2004) Robust sub-sentential alignment of phrase-structure trees. In Coling: 20th international conference on computational linguistics, Geneva, Switzerland, pp 183–190
Groves D, Way A (2005a) Hybrid example-based SMT: The best of both worlds? In ACL-2005 Workshop on building and using parallel texts, Ann Arbor, MI, pp 183–190
Groves D, Way A (2005b) Hybrid data-driven models of machine translation. Mach Translat 19:301–323
Hall RP (1989) Computational approaches to analogical reasoning: A comparative analysis. Artif Intell 39:39–120
Hutchins J (2005) Towards a definition of example-based machine translation. In: MT Summit X workshop: Second workshop on example-based machine translation, Phuket, Thailand, pp 63–70
Hutchins J (2005) Example-based machine translation: A review and commentary. Mach Translat 19:197–211
Hutchins WJ, Somers HL (1992) An introduction to machine translation. Academic Press, London, England
Kaji H, Kida Y, Morimoto Y (1992) Learning translation templates from bilingual text. In Proceedings of the fifteenth [sic] international conference on computational linguistics, COLING-92, Nantes, France, pp 672–678
Kibler D, Aha DW (1997) Learning representative exemplars of concepts: An initial study. In: Fourth international workshop on machine learning, Irvine, CA, pp 24–29
Kitano H, Higuchi T (1991) High performance memory-based translation on IXM2 massively parallel associative memory processor. In: AAAI-91: Proceedings of the ninth national conference on artificial intelligence, AAAI Press/ MIT Press, Menlo Park, CA, pp 149–154
Koehn P, Och FJ, Marcu D (2003) Statistical phrase-based translation. In: HLT-NAACL: Human language technology conference of the North American chapter of the Association for Computational Linguistics, Edmonton, Alberta, Canada, pp 127–133
Kolodner J (1983a) Maintaining organization in a dynamic long-term memory. Cog Sci 7:243–280
Kolodner J (1983b) Reconstructive memory, a computer model. Cog Sci 7:281–328
Langlais P, Gotti F (2005) EBMT by tree-phrasing. Mach Translat 20:1–25
Lepage Y, Denoual E (2005) Purest ever example-based machine translation: Detailed presentation and assessment. Mach Translat 19:251–282
Lin D, Cherry C (2003) Word alignment with cohesion constraint. In: HLT-NAACL 2003: Human language technology conference of the North American chapter of the Association for Computational Linguistics, Edmonton, Alberta, Canada, companion vol pp 49–51
Liu Z, Wang H, Wu H (2006) Example-based machine translation based on tree-string correspondence and statistical generation. Mach Translat 20:27–44
Locke WN, Booth AD (1955) Machine translation of languages: Fourteen essays. The Technology Press of the Massachusetts Institute of Technology, Cambridge, Mass/John Wiley, New York, NY/Chapman & Hall, London, England
Lu Y, Li S, Zhao T, Yang M (2002) Learning Chinese bracketing knowledge based on a bilingual language model. In: Proceedings of the 18th international conference on computational linguistics: COLING 2000 in Europe, Saarbrücken, Germany, pp 591–598
Lu Y, Zhou M, Li S, Huang C, Zhao T (2001) Automatic translation template acquisition based on bilingual structure alignment. Comput Ling Chin Lang Proc 6:83–108
Maas H-D (1987) The MT system SUSY. In King M (ed) Machine translation today: The state of the art. Edinburgh University Press, Edinburgh, Scotland, pp 209–246
Matsumoto Y, Ishimoto H, Utsuro T (1993) Structural matching of parallel texts. In: 31st annual meeting of the Association for Computational Linguistics, Columbus, OH, pp 23–30
McTait K, Trujillo A (1999) A language-neutral sparse-data algorithm for extracting translation patterns. In: Proceedings of the eighth international conference on theoretical and methodological issues in machine translation (TMI-99), Chester, England, pp 99–108
Merriam-Webster (2002) Webster's third new international dictionary. Merriam-Webster, Springfield, MA
Nagao M (1984) A framework of a mechanical translation between Japanese and English by analogy principle. In: Elithorn A, Banerji R (eds) Artificial and human intelligence (Edited review

papers presented at the international NATO symposium on artificial and human intelligence), North-Holland, Amsterdam, The Netherlands, pp 173–180; repr. in Nirenburg et al. (2003), pp 351–354
Nirenburg S, Somers H, Wilks Y (eds) Readings in machine translation. MIT Press, Cambridge, MA
Och FJ, Ney H (2002) Discriminative training and maximum entropy models for statistical machine translation. In: 40th annual meeting of the Association for Computational Linguistics, Philadelphia, PA, pp 295–302
Och FJ, Tillmann C, Ney H (1999) Improved alignment models for statistical machine translation. In: 1999 conference on empirical methods in natural language processing and very large corpora (EMNLPVLC-99), College Park, MD, pp 20–28
OED (2005) Compact Oxford English dictionary of current English. Oxford University Press, Oxford, England
Porter BW, Bareiss ER (1986) PROTOS: An experiment in knowledge acquisition for heuristic classification tasks. In: First international meeting on advances in learning (IMAL), Les Arcs, France, pp 159–174
Quirk C, Menezes A (2006) Dependency treelet translation: The convergence of statistical and example-based machine-translation? Mach Translat 20:45–66
Russell S, Wefald E (1991) Do the right thing: Studies in limited rationality. MIT Press, Cambridge, MA
Senellart J, Yang J, Rebollo A (2003) SYSTRAN intuitive coding technology. In: MT Summit IX: Proceedings of the ninth machine transaltion summit, New Orleans, USA, pp 346–353
Simard M, Langlais P (2003) Statistical translation alignment with compositionality constraints. In: HLT/NAACL-2003 Workshop on building and using parallel texts, Edmonton, Alberta, Canada, pp 19–22
Somers H (1999) Review article: Example-based machine translation. Mach Translat 14:113–157
Somers H (2003) An overview of EBMT. In: Carl and Way (2003), pp 3–57
Stanfill C, Waltz D (1988) The memory based reasoning paradigm. In: Case-based reasoning: Proceedings from a workshop, Morgan Kaufmann, San Mateo, CA, pp 414–424
Sumita E (2003) An example-based machine translation system using DP-matching between word sequences. In: Carl and Way (2003) pp 189–209
Sumita E, Iida H (1991) Experiments and prospects of example-based machine translation. In: 29th annual meeting of the Association for Computational Linguistics, Berkeley, CA, pp 185–192
Toma P (1976) An operational machine translation system. In: Brislin R (ed) Translation: Applications and research, Gardner Press, New York, NY, pp 247–259
Toma P (1977) Systran as a multilingual machine translation system. In: Overcoming the language barrier: Third European congress on information systems and networks, Luxembourg, May 1977, Verlag Dokumentation, München, pp 569–581
Turcato D, Popowich F (2003) What is example-based machine translation? In Carl and Way (2003), pp 59–81
Vauquois B (1975) La traduction automatique à Grenoble [Machine translation in Grenoble]. Dunod, Paris, France
Vauquois B, Boitet C (1985) Automated translation at Grenoble University. Comput Ling 11, 28–36
Veale T, Way A (1997) *Gaijin*: A bootstrapping, template-driven approach to example-based machine translation. In: International conference, recent advances in natural language processing, Tzigov Chark, Bulgaria, pp 239–244
Veloso MM, Carbonell JG (1993) Derivational analogy in Prodigy. Mach Learn 10:249–278
Vilar JM, Vidal E (2005) A recursive statistical translation model. In ACL-2005 Workshop on building and using parallel texts, Ann Arbor, MI, pp 199–207
Way A, Gough N (2005) Comparing example-based and statistical machine translation. Nat Lang Eng 11:295–309
Weaver W (1949) Translation. In: Locke and Booth (1955) pp 15–23; repr. in Nirenburg (2003), pp 13–17
WordNet (2003) WordNet 2.0. Cognitive Science Laboratory, Princeton University, Princeton, NJ
Wu D (1995a) An algorithm for simultaneously bracketing parallel texts by aligning words. In: 33rd annual meeting of the Association for Computational Linguistics, Cambridge, MA, pp 244–251
Wu D (1995b) Trainable coarse bilingual grammars for parallel text bracketing. In: 3rd annual workshop on very large corpora (WVLC-3), Cambridge, MA, pp 69–81
Wu D (1996) A polynomial-time algorithm for statistical machine translation. In: 34th annual meeting of the Association for Computational Linguistics, Santa Cruz, CA, pp 152–158

Wu D (1997) Stochastic inversion transduction grammars and bilingual parsing of parallel corpora. Comput Ling 23:377–404
Wu D (2000) Alignment. In: Dale R, Moisl H, Somers H (eds) Handbook of natural language processing, Marcel Dekker, New York, NY, pp 415–458
Wu D (2005) Recognizing paraphrases and textual entailment using inversion transduction grammars. In: ACL-2005 workshop on empirical modeling of semantic equivalence and entailment, Ann Arbor, MI, pp 25–30
Wu D, Fung P (2005) Inversion transduction grammar constraints for mining parallel sentences from quasi-comparable corpora. In: Second international joint conference on natural language processing (IJCNLP-2005), Jeju, Korea, pp 257–268
Wu D, Wong H (1998) Machine translation with a stochastic grammatical channel. In: COLING-ACL '98: 36th annual meeting of the Association for Computational Linguistics and 17th international conference on computational linguistics, Montreal, Quebec, Canada, pp 1408–1415
Yamada K, Knight K (2001) A syntax-based statistical translation model. In: Association for Computational Linguistics 39th annual meeting and 10th conference of the European chapter, Toulouse, France, pp 523–529
Yamamoto K, Matsumoto Y (2003) Extracting translation patterns from parallel corpora. In: Carl and Way (2003), pp 365–395
Zens R, Ney H (2003) A comparative study on reordering constraints in statistical machine translation. In: 41st annual meeting of the Association for Computational Linguistics, Sapporo, Japan, 192–202
Zens R, Ney H, Watanabe T, Sumita E (2004) Reordering constraints for phrase-based statistical machine translation. In: Coling: 20th international conference on computational linguistics, Geneva, Switzerland, pp 205–211
Zhang H, Gildea D (2005) Stochastic lexicalized inversion transduction grammar for alignment. In: 43rd annual meeting of the Association for Computational Linguistics, Ann Arbor, MI, pp 475–482
Zhang Y, Vogel S (2005) An efficient phrase-to-phrase alignment model for arbitrarily long phrase and large corpora. In: Proceedings of the 10th annual conference of the European Association for Machine Translation, Budapest, Hungary, pp 294–301
Zhao B, Vogel S (2003) Word alignment based on bilingual bracketing. In HLT/NAACL-2003 workshop on building and using parallel texts, Edmonton, Alberta, Canada, pp 15–18