20 MACHINE TRANSLATION*

> *. . . Translation is a fine and exacting art, but there is much about it that is mechanical and routine* Kay (1997)

MACHINE TRANSLATION
MT

This chapter introduces techniques for **machine translation** (**MT**), the use of computers to automate some or all of the process of translating from one language to another. Translation, in its full generality, is a difficult, fascinating, and intensely human endeavor, as rich as any other area of human creativity. Consider the following passage (with the Chinese transcribed in the Mandarin dialect) from the end of Chapter 45 of the David Hawkes translation of Cao Xue Qin's famous 18th-century Chinese novel *The Story of the Stone*, also called *Dream of the Red Chamber* (Cao, 1973)

> As she lay there alone, Dai-yu's thoughts turned to Bao-chai. . . Then she listened to the insistent rustle of the rain on the bamboos and plantains outside her window. The coldness penetrated the curtains of her bed. Almost without noticing it she had begun to cry.

dai yu	*zi*	*zai*	*chuang*	*shang*	*gan nian*	*bao chai. . .*
Dai-yu	alone	on	bed	top	think-of-with-gratitude	Bao-chai

you	*ting jian*	*chuang*	*wai*	*zhu*	*shao*	*xiang*	*ye*	*zhe*
again	listen to	window	outside	bamboo	tip	plantain	leaf	of

shang,	*yu*	*sheng*	*xi*	*li,*	*qing*	*han*	*tou*	*mu,*
on-top,	rain	sound	sigh	drip,	clear	cold	penetrate	curtain,

bu	*jue*	*you*	*di*	*xia*	*lei*	*lai.*
not	feeling	again	fall	down	tears	come.

Consider some of the issues involved in this kind of literary translation. First, how are Chinese names to be translated? Cao often used wordplay relying on the fact that Chinese names have a transparent meaning. Hawkes

[0] this chapter mostly by Nigel Ward

chose to keep the Chinese syllables for names of main characters, but to translate names of servants by their meanings (Aroma, Skybright). Chinese rarely marks verbal aspect or tense; Hawkes thus had to decide to translate Chinese *tou* as *penetrated*, rather than say *was penetrating* or *had penetrated*. Next, Hawkes inserted the possessive pronoun *her* to make *her window* more appropriate for the mood of a quiet bedroom scene than *the window*. Knowing that the English reader, not knowing about Chinese bed-curtains, would incorrectly assume *curtain* meant window curtain, Hawkes translated it as *curtains of her bed*. Finally, Chinese prose tends to use symmetric four-character phrases like *bamboo tip plantain leaf*. Hawkes decided that this would be awkward in English, and left it as *bamboos and plantains*.

Clearly creating a translation of this sort requires a deep and rich understanding of the source language and the input text, and a sophisticated, poetic, and creative ability with the target language. The problem of automatically producing a high-quality translation of an arbitrary text from one language to another is thus far too hard to automate completely. But certain simpler translation tasks can be addressed with current computational models. In particular, machine translation system often focus on (1) tasks for which a **rough translation** is adequate, (2) tasks where a human post-editor can improve MT output, and (3) tasks limited to small **sublanguage** domains in which fully automatic high quality translation is achievable.

Information acquisition on the Web is the kind of 'information pull' task where readers are be willing to settle for a very rough translation. Consider the extracts from a French web page and their translation:

> Nous sommes une association type Loi de 1901, et notre raison d'être est de practiquer, de promouvoir, de faire découvrir le Paintball, et le cas échéant de supporter nos équipes de compétition: . . . Si vous avez des questions, des envies d'organisation de parties, des envies de jouer tout courte et des envies de découvrir, n'hésitez pas à nous contacter par courrier ou par téléphone ou bien encore par eMail. . . . Au sortir de la saison 97/98 et surtout au début de cette saison 98/99, les effectifs des HORS-TAXE sont modifiés.

> We are a standard association Loi of 1901, and our raison d'ecirc;tre is to practice, promote, make discover Paintball, and to support our teams of competition if necessary: . . . If you have questions, desires of organization of parts, desires for playing very short and desires for discovering, do not hesitate to contact us by mail or telephone or even by eMail. . . . With leaving season 97/98 and especially at the beginning of this season 98/99, manpower

of the HORS-TAXE are modified!

This is good enough to figure out that we have the found the home page of a paintball team, and one that seems friendly and perhaps willing to accept new members. Armed with this information, we can then try to find someone to do a proper translation for us, or perhaps just go ahead and send e-mail to the organizer to ask if we can play. The use of MT for such document-finding purposes can sometimes be avoided or made more efficient by using **cross-language information retrieval** techniques, which focus on the retrieval of documents in a language other than that used for the query terms (Oard, 1997).

CROSS-LANGUAGE INFORMATION RETRIEVAL

Rough translation is also useful as the first stage in a complete translation process. An MT system can produce a draft translation that can be fixed up in a **post-editing** process by a human translator. Even a rough draft can sometimes speed up the overall translation process. Strictly speaking, systems used in this way are doing **computer-aided human translation** (CAHT or CAT) rather than (fully automatic) machine translation. This model of MT usage is effective especially for high volume jobs and those requiring quick turn-around. The most familiar example is perhaps the translation of software manuals for **localization** to reach new markets. Another effective application is the translation of market-moving financial news, for example from Japanese to English for use by stock traders.

POST-EDITING

COMPUTER-AIDED HUMAN TRANSLATION

LOCALIZATION

Weather forecasting is an example of a **sublanguage** domain that can be modeled completely enough to use raw MT output even without post-editing. Weather forecasts consist of phrases like *Cloudy with a chance of showers today and Thursday.*, *Low tonight 4, high Thursday 10.* and *Outlook for Friday: sunny.* This domain has a limited vocabulary and only a few basic phrase types. Ambiguity is rare, and the senses of ambiguous words are distinct and easily disambiguated based on local context, using word classes and semantic features such as MONTH, PLACE, DIRECTION, TIME POINT, TIME DURATION, DEGREE-OF-POSSIBILITY. Other domains that are sublanguage-like include equipment maintenance manuals, air travel information queries, appointment scheduling, and restaurant recommendations.

SUBLANGUAGE

This chapter breaks with the pattern of previous chapters in that the focus is less on introducing new techniques than on showing how the techniques presented earlier are used in practice. One of the themes of this chapter is that there are often trade-offs and difficult choices among alternative approaches and techniques.

Section 20.1 gives some simple illustrations of the ways in which languages differ. The following four sections then introduce four strategies for doing MT: Section 20.2 introduces the use of syntactic transformations for overcoming differences in grammar, as well as some techniques for choosing target language words. Section 20.3 introduces some ways of exploiting meaning during translation, in particular the use of thematic roles and primitive decomposition. 20.4 presents the minimalist direct approach. Section 20.5 discusses the use of statistical techniques to improve various aspects of MT. Finally, Section 20.6 discusses reasons for the gap between expectations and performance, and discusses strategies for meeting users' needs despite finite development resources.

20.1 Language Similarities and Differences

When you accidentally pick up a radio program in some foreign language it seems like chaos, completely unlike the familiar languages of your everyday life. But there are patterns in this chaos, and indeed, some aspects of human language seem to be **universal**, holding true for every language. Many universals arise from the functional role of language as a communicative system by humans. Every language, for example, seems to have words for referring to people, for talking about men, women, and children, eating and drinking, for being polite or not. Other universals are more subtle; for example Chapter 8 mentioned that every language seems to have nouns and verbs.

UNIVERSAL

Even when languages differ, these differences often have systematic structure. The study of systematic cross-linguistic similarities and differences is called **typology** (Croft (1990), Comrie (1989)). This section sketches some typological facts about crosslinguistic similarity and difference. This bears on our main topic, MT, in that the difficulty of translating from one language to another depends a great deal on how similar the languages are in their vocabulary, grammar, and conceptual structure.

TYPOLOGY

Morphologically, languages are often characterized along two dimensions of variation. The first is the number of morphemes per word, ranging from **isolating** languages like Vietnamese and Cantonese, in which each word generally has one morpheme, to **polysynthetic** languages like Siberian Yupik (Eskimo), in which a single word may have very many morphemes, in effect corresponding to a whole sentence in English. The second dimension is the degree to which morphemes are segmentable, ranging from **agglutina-**

ISOLATING

POLYSYNTHETIC

tive languages like Turkish (discussed in Chapter 3), in which morphemes have relatively clean boundaries, to **fusion** languages like Russian, in which a single affix may conflate multiple morphemes (like *-om* in the word *stolom*, (table-SG-INSTR-DECL1) which fuses distinct morphological categories instrumental, singular, and first declension.)

AGGLUTINATIVE

FUSION

Syntactically, languages are most saliently different in the basic word order of verbs, subjects, and objects in basic declarative clauses. German, French, English, and Mandarin are similar, for example, in being **SVO** languages, meaning that the verb tends to come between the subject and object in basic declarative clauses. Hindi and Japanese, by contrast, or **SOV** languages, meaning that the verb tends to come at the end of basic clauses, while Irish, Classical Arabic, and Biblical Hebrew are **VSO** languages. Two languages that share their basic word-order type often have other similarities. For example **SVO** languages generally have **prepositions** while **SOV** languages generally have **postpositions**. Thus the English PP *to Yuriko* would appear in Japanese as *Yuriko ni*.

SVO

SOV

VSO

Another important syntactico-morphological distinction is between **head-marking** and **dependent-marking** languages (Nichols, 1986). Head-marking languages tend to mark the relation between the head verb of a clause and its dependents (or the head noun of a noun-phrase and its dependents) on the head. Dependent-marking languages tend to mark the relation on the non-head. Nichols (1986), for example, notes that Hungarian marks the possessive relation with an affix (A) on the head noun (H), where English marks it on the (non-head) possessor:

HEAD-MARKING

DEPENDENT-MARKING

(20.1)	*English*	the	man-^{A}'s	^{H}house
	Hungarian	az	ember	^{H}ház-^{A}a
		the	man	house-his

This syntactic distinction is related to a semantic distinction in how languages map conceptual notions onto words. Talmy (1985) and (1991) noted that languages can be characterized by whether **direction of motion** and **manner of motion** are marked on the verb or on what Talmy called the **satellite**; particles, prepositional phrases, or adverbial phrases. For example a bottle floating out of a cave would be described in English with the direction marked on the particle *out* as:

(20.2) The bottle floated out.

but in Spanish with the direction marked on the verb as

(20.3) *La botella salió flotando.*
The bottle exited floating.

Languages that mark the direction of motion on the verb (leaving the satellites to mark the manner of motion) Talmy called verb-framed; Slobin (1996) gives examples like Spanish *acercarse* 'approach', *alcanzar* 'reach', *entrar* 'enter', *salir* 'exit'. Languages that mark the direction of motion on the satellite (leaving the verb to mark the manner of motion) Talmy called satellite-framed; Slobin (1996) gives examples like English *crawl out*, *float off*, *jump down*, *walk over to*, *run after*. Talmy (1991) noted that verb-framed languages include Romance, Semitic, Japanese, Tamil, Polynesian, most Bantu, most Mayan, Nez Perce, and Caddo, while satellite-framed languages include most Indo-European minus Romance, Finno-Ugric, Chinese, Ojibwa, and Warlpiri.

In addition to such properties that systematically vary across large classes of languages, there are many specific characteristics, more or less unique to single languages. English, for example, has an idiosyncratic syntactic construction involving the word *there* that is often used to introduce a new scene in a story, as in *there burst into the room three men with guns*.

To give an idea of how trivial, yet crucial, these differences can be, think of dates. Dates not only appear in various formats — YYMMDD in Japanese, MM-DD-YY in American English, and DD-MM-YY in British English — the calendars themselves may differ, for example dates in Japanese often are relative to the start of the current Emperor's reign rather than to the start of the Christian Era.

Turning now to the question of lexical organization, here too there are interesting patterns. Many words can be translated relatively directly into other languages. English *dog*, for example, translates to Mandarin *gou*. Where English has *chocolate*, Italian has *cioccolato* and Japanese has *chokoreeto*.[1]

Sometimes, rather than a single word, there is a fixed phrase in the target language; French *informatique* thus translates to English *computer science*. In more difficult cases, however, a word in one language does not map so simply to a word or phrase in another language. This can be caused by a difference in the **grammatical** constraints imposed by the source and target language, or a difference in how the two languages lexically break up their **conceptual** space.

[1] although *chokoreeto* in Japanese is perforce more formal than English *chocolate*, since Japanese also has the informal short form *choko*.

Grammatically, for example, a word may translate best to a word of another part of speech in the target language. Many English sentences involving the verb *like* must be translated into German using the adverbial *gern*; thus *she likes to sing* maps to *sie singt gerne*, where the syntactic structure is also affected.

Sometimes one language places more grammatical constraints on word choice than another. English, for example, distinguishes gender in pronouns where Mandarin does not; Thus translating a third-person singular pronoun from Mandarin to English requires deciding whether the original referent was masculine or feminine. The same is true when translating from the English pronoun plural *they*, unspecified for gender, into French (masculine *ils*, feminine *elles*). In Japanese, there is no single word for *is*, speakers must choose between *iru* or *aru*, based on whether the subject is animate[2] or not.

This kind of difference in specificity also happens in the conceptual domain; one language may divide up a particular conceptual domain in more detail than another. English, for example, has a particularly impoverished kinship vocabulary; the single word *brother* can indicate either a younger or older brother. Both Japanese and Chinese, by contrast, both distinguish seniority in sibling relations. Figure 20.1 gives some further examples.

The way that languages differ in lexically dividing up conceptual space may be more complex than this one-to-many translation problem, leading to many-to-many mappings. For example Figure 20.2 summarizes some of the complexities cited by (Hutchins and Somers, 1992) in relating English *leg, foot*, and *paw*, and French *jambe, pied, patte*, etc.

Further, one language may have a **lexical gap**, where no word or phrase, short of an explanatory footnote, can express the meaning of a word in the other language. For example, Japanese does not have a word for *privacy*, and English does not have a word for Japanese *oyakoko* (we make do with the Latin *filial piety*). LEXICAL GAP

Moreover, dependencies on cultural context, as manifest in the background and expectations of the readers of the original and translation, further complicate matters. A number of translation theorists (Steiner, 1975; Barnstone, 1993; Hofstadter, 1997) refer to a clever story by Jorge Luis Borges showing that even two linguistic texts with the same words and grammar may have different meanings because of their different cultural contexts. Borges creates Menard, a French author in the 1930's whose aim was to recreate Cervantes' *Don Quixote* word for word:

[2] Taxis and buses in service sometimes count as animate for this purpose.

English	*brother*	Japanese	*otooto* (younger)
		Japanese	*oniisan* (older)
		Mandarin	*gege* (older)
		Mandarin	*didi* (older)
English	*wall*	German	*Wand* (inside)
		German	*Mauer* (outside)
English	*know*	French	*connaître* (be acquainted with)
		French	*savoir* (know a proposition)
English	*they*	French	*ils* (masc)
		French	*elles* (fem)
German	*berg*	English	*hill*
		English	*mountain*
Mandarin	*tā*	English	*he*
		English	*she*
		English	*it*

Figure 20.1 Differences in conceptual specificity. Translating from a more specific to a less specific language, a phrase can often be used (e.g. English *younger brother* for Japanese *otooto*). It is much harder to be faithful when translating from a less specific to a more specific language.

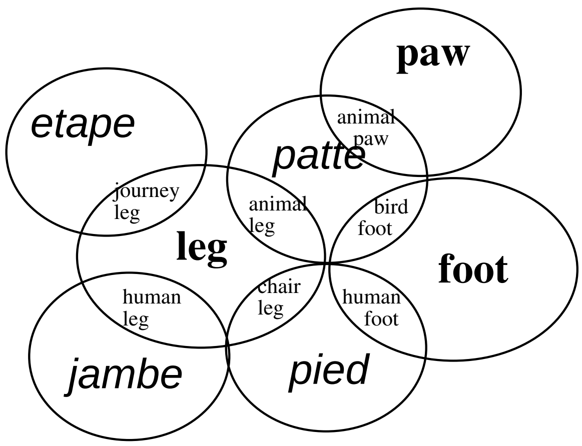


Figure 20.2 The complex overlap between English *leg*, *foot*, etc, and various French translations.

The text of Cervantes and that of Menard are verbally identical, but the second is almost infinitely richer. (More ambiguous, his detractors will say; but ambiguity is a richness.) It is a revelation to compare the *Don Quijote* of

> Menard with that of Cervantes. The latter, for instance, wrote:
>
> > *... la verdad, cuya madre es la historia, émula del tiempo, depósito de las acciones, testigo de lo pasado, ejemplo y aviso de lo presente, advertencia de lo por venir.*
>
> Menard, on the other hand, writes:
>
> > *... la verdad, cuya madre es la historia, émula del tiempo, depósito de las acciones, testigo de lo pasado, ejemplo y aviso de lo presente, advertencia de lo por venir.*
>
> Equally vivid is the contrast in styles. The archaic style of Menard – in the last analysis, a foreigner — suffers from a certain affectation. Not so that of his precursor, who handles easily the ordinary Spanish of his time.

These last points suggest a more general question about cultural differences and the possibility (or impossibility) of translation. A theoretical position sometimes known as the **Sapir-Whorf hypothesis** suggests that language may constrain thought — that the language you speak may affect the way you think. To the extent that this hypothesis is true, there can be no perfect translation, since speakers of the source and target languages necessarily have different conceptual systems. In any case it is clear that the differences between languages run deep, and that the process of even partial translation is not going to be simple.

SAPIR-WHORF HYPOTHESIS

20.2 THE TRANSFER METAPHOR

As the previous section illustrated, languages differ. One strategy for doing MT is to translate by a process of overcoming these differences, altering the structure of the input to make it conform to the rules of the target language. This can be done by applying **contrastive** knowledge, that is, knowledge about the differences between the two languages. Systems that use this strategy are sometimes said to be based on the **transfer model**.

CONTRASTIVE

TRANSFER MODEL

Since this requires some representation of the structure of the input, transfer presupposes a parse of some form. Moreover, since transfer only results in a structure for the target language, it must be followed by a generation phase to actually create the output sentence. Thus, on this model, MT involves three phases: **analysis**, **transfer**, and **generation**, where transfer bridges the gap between the output of the source language parser and the input to the target language generator. Figure 20.3 shows a sketch of this transfer architecture.

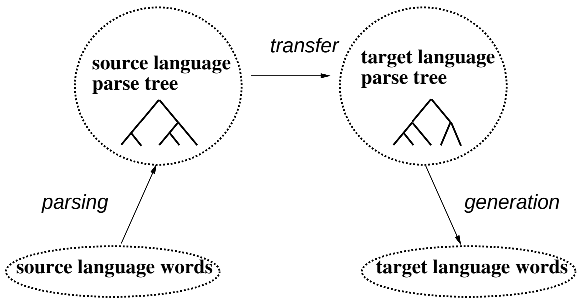


Figure 20.3 The transfer architecture for Machine Translation.

It is worth noting that a parse for MT may differ from parses required for other purposes. For example, suppose we need to translate *John saw the girl with the binoculars* into French. The parser does not need to bother to figure out where the prepositional phrase attaches, because both possibilities lead to the same French sentence.

Syntactic Transformations

Let us begin by considering syntactic differences. The previous section noted that in English the unmarked order in a noun-phrase had adjectives precede nouns, but in French adjectives follow nouns.[3] Temporarily postponing the question of how to translate the words, let's consider how an MT system can overcome such differences.

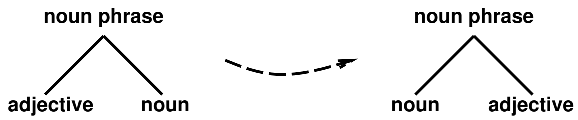


Figure 20.4 A simple transformation that reorders adjectives and nouns

SYNTACTIC TRANSFORMATIONS

Figure 20.4 suggests the basic idea. Here we transform one parse tree, suitable for describing an English phrase, into another parse tree, suitable for describing a French sentence. In general, **syntactic transformations** are

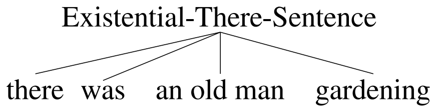
[3] There are exceptions to this generalization, such as *galore* in English and *gros* in French; furthermore in French some adjectives can appear before the noun with a different meaning; *route mauvaise* 'bad road, badly-paved road' versus *mauvaise route* 'wrong road' (Waugh, 1976).

operations that map from one tree structure to another.

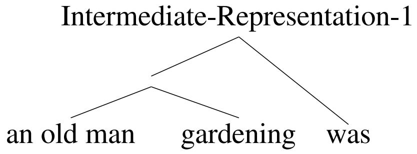
Now let's illustrate how roughly how such transformations can restructure an entire sentence. In order to show the mechanism in detail, we'll use a simplified sentence:

(20.4) There was an old man gardening.

We will assume that the parser has given us a structure like the following. We will also assume that the system starts performing transformations at the top node of the tree and works its way down:

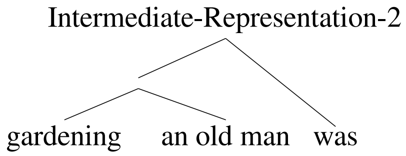


Since this sentence involves an "existential *there* construction", which has no analog in Japanese, we immediately have to apply a transformation that deletes the sentence-initial *there* and converts the fourth constituent to a relative clause modifying the noun, producing something like following structure:

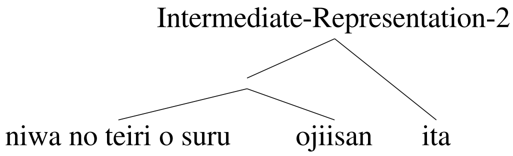


The resulting structure is thus something more like the structure of *an old man, who was gardening, was*, which is already beginning to have the verb-final structure of Japanese.

Next, another transformation applies to reverse the order of the noun phrase and the relative clause, giving something like the following structure:



At this point all relevant transformations have applied, and lexical transfer takes place, substituting Japanese words for the English ones, as discussed in the next section. This gives the final structure below:



After this, a little more syntactic work is required to produce an actual Japanese sentence, including: 1. adding the word *ga*, which is required in Japanese to mark the subject, 2. choosing the verb that agrees with the subject in terms of animacy, namely *iru*, not *aru*, and 3. inflecting the verbs. The final generation step traverses or otherwise linearizes the tree to produce a string of words. Although these generation tasks can be done by the techniques of Chapter 19, practical systems usually do them directly with simple procedures. In any case, the final output will be:

niwa	*no*	*teiri*	*o*	*shite*	*ita*	*ojiisan*	*ga*	*ita.*
garden	GEN	upkeep	OBJ	do	PAST-PROG	old man	SUBJ	was

Table 20.5 shows a rough representation of the transformations we have discussed. Such transformations can be implemented as pattern-rewrite rules: if the input matches the left side of a transformation, it is rewritten according to the right side.

English to French:	
1.	NP $\rightarrow$ Adjective$_1$ Noun$_2$ $\Rightarrow$ NP $\rightarrow$ Noun$_2$ Adjective$_1$
Japanese to English:	
2.	Existential-There-Sentence $\rightarrow$ There$_1$ Verb$_2$ NP$_3$ Postnominal$_4$ $\Rightarrow$ Sentence $\rightarrow$ (NP $\rightarrow$ NP$_3$ Relative-Clause$_4$) Verb$_2$
3.	NP $\rightarrow$ NP$_1$ Relative Clause$_2$ $\Rightarrow$ NP $\rightarrow$ Relative-Clause$_2$ NP$_1$

Figure 20.5 An informal description of some transformations.

Transformations in MT systems also may include more complex conditions for when they apply, and may have a "trigger", that is, a specific word that is used to index the pattern, for efficiency. One way to formalize transformations is with unification-based models; indeed as Chapter 11 discussed, the need for a reversable operation for MT was the original motivation for both feature-structure unification (Kay, 1984) and term-unification (Colmerauer and Roussel, 1996). However, unification, despite its advantages, is computationally expensive and hence is not very commonly used.

Lexical Transfer

Some of the output words are determined in the course of syntactic transfer or generation. In the example above, the function words *ga* and *ita* are mostly grammatically controlled. Content words are another matter. The process of finding target language equivalents for the content words of the input is called **lexical transfer**. The problems that arise in lexical transfer were introduced in Section 20.1. They are generally addressed using techniques built on the methods described in earlier chapters. LEXICAL TRANSFER

The foundation of lexical transfer is dictionary lookup in a crosslanguage dictionary. As was discussed earlier, the translation equivalent may be a single word or a it may be a phrase (as in this example where *gardening* becomes *niwa no teiri o suru* ('do garden upkeep')). Furthermore, sometimes a generation process must subsequently inflect words in such phrases, as in this case.

Section 20.1 also discussed the problem of words that have several possible translations. In the example *man* is such a word. The correct choice here was *ojiisan* ('old man'), but if the input had been *man is the only linguistic animal*, the translation of *man* would have been *ningen* ('human being, man, men'); in most other cases *hito* ('person, persons, man, men') or related words would have been appropriate. Fortunately there are at least two ways to tackle this problem: in the parsing or in the generation stage. The first method is to treat words like *man* as if they were ambiguous. That is, we assume that *man* can correspond to two more more concepts (perhaps HUMAN and ADULT MALE) and that choosing the correct Japanese word is like disambiguating between these concepts. This way of treating lexical transfer lets us apply all the standard techniques for lexical disambiguation in Chapter 16. A second way is to treat *man* as having only one meaning, and to treat the multiple possible translations (*ningen*, *hito*, *ojiisan* and so on) as caused by constraints imposed by the target language in generation. In practice, these cases are more often dealt with in the parsing stage, as the algorithms for choosing content words during generation are high-overhead (Ward, 1994).

In this specific example, however, the choice of how to translate *man* is easy. Because the previous word is *old*, the correct translation is *ojiisan* ('old man'). Such inputs, where multiple source language words must be expressed with a single target language word, can be difficult to handle, requiring inference in the general case. But many such cases, including this one, can be treated simply as idioms, with their own entries in the bilingual dictionary.

20.3 THE INTERLINGUA IDEA: USING MEANING

One problem with the transfer model is that it requires a distinct set of transfer rules for each pair of languages. This is clearly suboptimal for translation systems employed in multilingual environments like the European Union, where (at least) seven official languages need to be intertranslated.

One approach to avoiding this requires a change in perspective regarding the nature of translation. The transfer model treats translation as a process of altering the structure and words of an input sentence to arrive at a valid sentence of the target language. An alternative to is to treat translation as a process of extracting the meaning of the input and then expressing that meaning in the target language. If this can be done, a MT system can do without contrastive knowledge, merely relying on the same syntactic and semantic rules used by a standard interpreter and generator for the language. The amount of knowledge needed is then proportional to the number of languages the system handles, rather than to the square.

INTERLINGUA

In order for this process to work for any pair of languages, the meaning representation, or **interlingua**, must be in a language-independent canonical form like the semantic representations we saw in Chapter 14. The idea is for the interlingua to represent all sentences that mean the 'same' thing in the same way, regardless of the language they happen to be in. Translation in this model proceeds by performing a semantic analysis on the input from language X into the interlingual representation and generating from the interlingua to language Y.

A frequently used element in interlingual representations is the notion of a small fixed set of thematic roles, as discussed in Chapter 16. When used in an interlingua, these thematic roles are taken to be language universals. Figure 20.6 shows a possible interlingual representation for *there was an old man gardening* as a unification-style feature structure that combines semantic and syntactic information. [4] We saw in Chapter 15 how a **semantic analyzer** can produce such a structure with a AGENT relation between *man* and *gardening*. Note that since the interlingua requires semantic interpre-

[4] Of course this is seriously inadequate as an account of the meaning of the existential-there construction. In fact, the currently least incomplete account of the syntax and semantics of *there* constructions in English takes 124 pages (Lakoff, 1987).

$$\begin{bmatrix} \text{EVENT} & \text{GARDENING} \\ \text{AGENT} & \begin{bmatrix} \text{MAN} & \\ \text{NUMBER} & \text{SG} \\ \text{DEFINITENESS} & \text{INDEF} \end{bmatrix} \\ \text{ASPECT} & \text{PROGRESSIVE} \\ \text{TENSE} & \text{PAST} \end{bmatrix}$$

Figure 20.6 Interlingual representation of *there was an old man gardening*.

tation in addition to syntactic parsing, it requires more analysis work than the transfer model, which only required syntactic parsing. But generation can now proceed directly from the interlingua with no need for syntactic transformations.

Note that the representation shown in Figure 20.6 includes the value GARDENING as the value for the EVENT feature. While this use of GARDENING is consistent with the event representations described in Chapter 14, it looks quite a bit more like an English word than an element in an interlingua. As it is being used here, GARDENING is intended to represent a language independent element in the **ontology** of this interlingua. That is, it is an element drawn from the inventory of concepts and relations that constitute the range of things expressible with this interlingua.

ONTOLOGY

Unfortunately, ontology design is more of an art than a science, and to make matters worse, the choice of ontology turns out to constrain the architecture of the entire MT system. For example, recall from Chapter 16 the discussion of two possible inventories of thematic roles, one containing AGENT and FORCE, and one including AGENT only. The choice of which to adopt affects, for example, the way that the system will translate *the quake broke glass* (Chapter 16) into Japanese, where *quake* needs to be marked with *de*, not the usual subject marker *ga*, because the earthquake is not animate. If we design our interlingua using the smaller inventory that only uses AGENT, then the representation for this sentence will place the *quake* in the AGENT role, and the problem of *de* versus *ga* will fall to the generator. If, however, we use the inventory shown in Table **??**, then the representation will include FORCE, with the work needed to make that decision being performed by the semantic analyzer.

The interlingua idea has implications not only for syntactic transfer but also for lexical transfer. The idea is to avoid explicit descriptions of the relations between source language words and target language words, in favor of mapping via concepts, that is, language-independent elements of the ontology. Recalling our earlier problem of whether to translate *man* as *otoko*, *ningen*, *ojiisan*, etc. it is clear that most of the processing involved is not specific to the goal of translating into Japanese; there is a more general problem of disambiguating *man* into concepts such as GENERIC-HUMAN and MALE-HUMAN. If we commit to using such concepts in an interlingua, then a larger part of the translation process can be done with general language processing techniques and modules, and the processing specific to the English-to-Japanese translation task can be eliminated or at least reduced.

Some interlinguas, and some other representations, go further and use lexical decomposition, that is, the disassembly of words into their component meanings. We saw a form of this in Figure 20.6, where *was* maps to PAST and PROGRESSIVE, and *a* maps to SINGULAR and INDEFINITE. Decomposition of content words is also possible: the word *drink* can be represented by (INGEST, FLUID, BY-MOUTH)[5]. Representing a sentence by breaking down the words in such ways does seem to be actually capturing something about meaning, rather than being just a rearrangement of the lexical items of the input. Moreover, such representations are potentially useful for inference-based disambiguation. For example, it is possible to use the meanings of the words to infer what the prepositional phrase is modifying in *the policeman saw the man with a telescope*, versus *the policeman shot the man with a telescope.* It is, however, difficult to get inference of this sort to work for more than a few examples except in very small domains. In general, such high-powered interlingua-based techniques are not used in practice.

Brushing over innumerable important details, we can now contrast the transfer model with the interlingua model. The key implication for processing is that, by making the parser/interpreter and/or the generator do a little more work, we can eliminate the need for contrastive knowledge, as suggested in Figure 20.7.

Doing the extra work involved by the interlingua commitment, however, is not always easy. It requires the system designer to perform exhaustive analysis of the semantics of the domain, that is, the formalization of what sorts of entities exist and in what possible ways they can relate to each other. This is possible in domains where the topic of discourse can, in essence, be expressed in what we earlier referred to as a sublanguage. Perhaps, not sur-

[5] This use of semantic decomposition makes it clear which elements of meaning *drink* shares with *eat* and which it does not share. But as Chapter 16 discusses, lexical semantics is not so easy in general. For example, how does one express in a formal language the meaning of *heft* and the way it differs from *weight*, or the meanings of *sporadic* and *intermittent*?

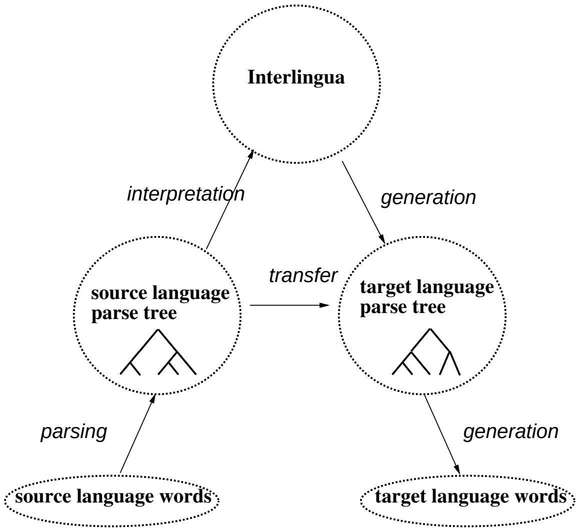


Figure 20.7 Diagram Suggesting the Relation Between the Transfer and Interlingua Models, after ? (?)

prisingly the semantics of these domains are often captured by a pre-existing database model, as in the air travel, hotel reservation, or restaurant recommendation domains. In cases like these, the database definition determines the possible entities and relations; and the MT designer just has to determine how these map to the words and structures of the two languages. Even this effort is not trivial (Levin *et al.*, 1998).

Another problem with the interlingua idea is that, in its pure form, it requires the system to fully disambiguate at all times. For a true universal interlingua, this may require some unnecessary work. For example, in order to translate from Japanese to Chinese the interlingua must include concepts such as ELDER-BROTHER and YOUNGER-BROTHER. However, to use those same concepts in the course of translating from German-to-English would require a parser to perform more disambiguation effort than is unnecessary; and will further require the system to make an extra effort to **preserve ambiguity**, to ensure that the output is ambiguous in exactly the same way as the input. Even discounting the Sapir-Whorf idea, the idea of a universal meaning underlying all languages is clearly not without problems. PRESERVE AMBIGUITY

20.4 DIRECT TRANSLATION

These models are all very nice, but what happens if the analysis fails? Users do not like to receive an output of "nil" due to "no parse tree found"; in general, they would rather get something imperfect than nothing at all. This is a challenge especially for interlingua-based models, where the system should not fail to translate *it broke the glass* because it can not figure out whether *it* is a FORCE or AGENT.

Several approaches are available. One is to use the robust parsing techniques discussed in Chapter 15, which sometimes amounts to translating by fragments. Another is to give up on producing elaborate structural analyses at all, and just do simple operations that can be done reliably. More radically, we could adopt the principle that a MT system should do as little work as possible. Systems built according to this philosophy are sometimes called

DIRECT

direct MT systems. Typically such systems are built with only one language pair in mind, and the only processing done is that needed to get from one specific source language to one specific target language.

A direct MT system is typically composed of several stages, each focused on one such problem. For example, we can rewrite a Japanese to English sentence in six stages, as seen in Figure 20.8. Figure 20.9 illustrates

Stage	Action
1.	morphological analysis
2.	lexical transfer of content words
3.	various work relating to prepositions
4.	SVO rearrangements
5.	miscellany
6.	morphological generation

Figure 20.8 Six Stages for a Direct MT System for Japanese to English

how this might work for a simple example.

Stage 1 in Figure 20.9 segments the input string into words (recall that Japanese, like Chinese, does not use spaces as word boundary markers), and does morphological analysis of complex verb forms. These can be done using the finite-state techniques of Chapter 3 and segmentation algorithms like the probabilistic one described in Chapter 5.

Stage 2 chooses translation equivalents for the content words. This is done using a bilingual dictionary, or procedures that choose the correct trans-

Input:	watashihatsukuenouenopenwojonniageta.
After stage 1:	watashi ha tsukue no ue no pen wo jon ni ageru PAST.
After stage 2:	I ha desk no ue no pen wo John ni give PAST.
After stage 3:	I ha pen on desk wo John to give PAST.
After stage 4:	I give PAST pen on desk John to.
After stage 5:	I give PAST the pen on the desk to John.
After stage 6:	I gave the pen on the desk to John.

Figure 20.9 An Example of Processing in a Direct System

lation based on the local context and on the target language words already chosen. Figure 20.10 illustrates such a procedure.

In this example lexical transfer is trivial. In general, though, there may be interdependencies among target-language words, and so lexical transfer this may be done in sub-stages, for example, verbs before nouns before adjectives. For example, consider the problem of translating *nomu* from Japanese to English, where this must become either *drink* or *take* (medicine). This decision must be made before translations for modifiers are chosen, to allow translations such as *drinking heavily* and *taking a lot of medicine*, but not a scramble of the two. In general the problem of the best order in which to make decisions is a tricky one, although there are some standard solutions, as seen in Chapter 19.

Stage 3 chooses to translate *no ue no* ('at top of') to *on*, and reverses the two associated noun phrases (*desk* and *pen*), since English prepositional phrases follow, not precede, the word they modify. In accordance with the dictionary entry for *gave*, which specifies subcategorization facts, it chooses to translate *ni* as *to*.

Stage 4 invokes a procedure to move the verb from the end of the sentence to the position after the subject, and removes case marking from direct and indirect objects.

Stage 5 handles things like moving prepositions before nouns and inserting articles.

Finally Stage 6 inflects the verbs.

There are several ways in which this approach differs from the approaches seen earlier. One is that it is a new way of modularizing the MT task, orthogonal to the types of modularity seen in the transfer and interlingua models in Figure 20.7. In the direct model, all the processing involving analysis of one specific issue (prepositions for example) is handled in one stage, including analysis, transfer, and generation aspects. The advantage of

```
function DIRECTLY_TRANSLATE_MUCH/MANY(Russian word) returns

if preceding word is how:
  return skol'ko
else if preceding word is as
  return stol'ko zhe
else if word is much
  if preceding word is very
    return nil (not translated)
  else if following word is a noun
    return mnogo
else /* word is many */
  if preceding word is a preposition and following word is a noun
  return mnogii
else return mnogo
```

Figure 20.10 A procedure for translating *much* and *many* into Russian, adapted from Hutchin's (1986, pg. 133) discussion of Panov 1960.

this is that solving specific problems one at a time may be more tractable. On the other hand, it is advantageous to organize processing into larger modules (analysis, transfer, synthesis) if there is synergy among all the various individual analysis problems, or among all the individual generation problems, etc.

A second characteristic of direct systems is that lexical transfer may be more procedural. Lexical transfer procedures may eclectically look at the syntactic classes and semantic properties of neighboring words and dependents and heads, as seen in the the decision-tree-like procedure for translating *much* and *many* into Russian in Figure 20.10.

A third characteristic of direct models is that they tend to be conservative, to only reorder words when required by obvious ungrammaticality in the result of direct word-for-word substitution. In particular, direct systems generally do lexical transfer before syntactic processing.

Perhaps the key characteristic of direct models is that they do without complex structures and representations. In general, they treat the input as a string of words (or morphemes), and perform various operations directly on it — replacing source language words with target language words, reordering words, etc. — to end up with a string of symbols in the target language. As such, the quality of such systems is almost entirely a function of the bilingual dictionary; and the larger the dictionary the better. Coverage of idioms, fixed phrases, and even common clauses and sentences is possible, especially if such data is extracted automatically from corpora. In some situations it may be valuable to do this on-line, at translation time, rather than saving the results in a dictionary — this is they key idea behind **Example-based Machine Translation** (Sumita and Iida, 1991; Brown, 1996).

EXAMPLE-BASED MACHINE TRANSLATION

In practice, of course, working MT systems tend to be combinations of the direct, transfer, and interlingua methods. But of course syntactic processing is not an all-or-nothing thing. Even if the system does not do a full parse, it can adorn its input with various useful syntactic information, such as part of speech tags, segmentation into clauses or phrases, dependency links, and bracketings. Many systems that are often characterized as direct translation systems also have aspects of the transfer and interlingua approaches (Hutchins and Somers, 1992).

20.5 Using Statistical Techniques

The three architectures for MT introduced in previous sections, the transfer, interlingua, and direct models, all provide answers to the questions of what representations to use and what steps to perform to translate. But there is another way to approach the problem of translation: to focus on the result, not the process. Taking this perspective, let's consider what it means for a sentence to be a translation of some other sentence.

This is an issue to which philosophers of translation have given a lot of thought. The consensus seems to be, sadly, that it is impossible for a sentence in one language to be a translation of a sentence in other, strictly speaking. For example, one cannot really translate Hebrew *HEBREW HERE* ('the Lord is my shepherd') into the language of a culture that has no sheep. On the one hand, we can write something that is clear in the target language, at some cost in fidelity to the original, something like *the Lord will look after me*. On the other hand, we can be faithful to the original, at the cost of producing something obscure to the target language readers, perhaps like *the Lord is for me like somebody who looks after animals with cotton-like hair*. As another example, if we translate the Japanese phrase *fukaku hansei shite orimasu*, as *we apologize*, we are not being faithful to the meaning of the original, but if we produce *we are (withdrawing from our usual activities and) deeply reflecting (on our past behavior, and what we did wrong, and how to avoid* *the problem next time)*, then our output is unclear or awkward. Problems such as these arise not only for culture-specific concepts, but whenever one language uses a metaphor, a tense, a construction, or a word without an exact parallel in the other language.

So, true translation, which is both faithful to the source language and natural as an utterance in the target language, is sometimes impossible. If you are going to go ahead and produce a translation anyway, you have to compromise. This is exactly what translators do in practice: they produce translations that do tolerably well on both criteria.

This provides us with a hint on how to do MT. We can model the goal of translation as the production of an output that maximizes some value function that represents the importance of both faithfulness and fluency. If we chose the product of fluency and faithfulness as our quality metric, we can formalize the translation problem as:

$$\text{best-translation } \hat{T} = \text{argmax}_T \text{ fluency(T) faithfulness(T,S)}$$

where T is the target-language-sentence and S the source-language-sentence.

This model of translation was first described by IBM researchers influenced by their own work on speech recognition (Brown *et al.*, 1990, 1993), and indeed the model clearly resembles the Bayesian models we've used for speech recognition in Chapter 7 and for spell checking in Section **??**. We can make the analogy perfect and apply the noisy channel model of Section **??** if we think of things backwards: thus we pretend that the input we must translate is a corrupted version of some target language sentence, and that our task is to discover that target language sentence:

$$\text{best-translation } \hat{T} = \text{argmax}_T \text{ P(T) } P(S|T)$$

To implement this, we need to do three things: quantify fluency, quantify faithfulness, and create an algorithm to find the sentence that maximizes the product of these two things.

There is an innovation here. In the transfer, interlingua, and direct models, each step of the process made some adjustment to the input sentence to make it closer to a fluent TL sentence, while obeying the constraint of not changing the meaning too much. In those models the process is fixed, in that there is no flexibility to trade-off a modicum of faithfulness for a smidgeon of naturalness, or conversely, based on the specific input sentence at hand. This new model, sometimes called the *statistical model* allows exactly that.

Quantifying Fluency

Fortunately, we already have some useful metrics for how likely a sentence is to be a real English sentence: the language models from Chapters 6 and 8. These allow us to distinguish things that are readable but not really English (such as *that car was almost crash onto me*) from things that are more fluent (*that car almost hit me*). This is especially valuable for word order and collocations, and as such can be a useful supplement to the generation techniques of Chapter 19.

Fluency models can be arbitrarily sophisticated; any technique that can assign a better probability to a target language string is appropriate, including the more sophisticated probabilistic grammars of Chapter 12 or the statistical semantic techniques of Chapter 16.

Of course, the idea of using monolingual language knowledge to improve MT output, is independent of the decision to model that knowledge statistically. Indeed, many MT systems, especially direct ones, have a final phase, in which the system uses local considerations of the output to revise word choices. For example, capitalizing every occurrence of *white house* that occurs as the subject of a verb (*the white house announced today*) is a reasonable heuristic.

Quantifying Faithfulness

Given the French sentence *ca me plaît* (*that me pleases*) and some conceivable English equivalents *that pleases me*, *I like it*, and *I'll take that one*, and *yes, good*, it is intuitively clear that the first is more faithful.

Although it is hard to quantify this intuition, one basic factor often used in metrics for fidelity is the degree to which the words in one are plausible translations of the words of the other. We can approximate the probability of a sentence being a good translation as the product of the probabilities that each target language word is an appropriate translation of some source language word. For this we need to know, for every source language word, the probability of it mapping to each possible target language word.

Where do we get these probabilities? Standard bilingual dictionaries do not include such information, but they can be computed from bilingual corpora, that is, parallel texts in two languages. This is not trivial, since bilingual corpora do not come with annotations specifying which word maps to which. Solving this problem requires first solving the problem of **sentence alignment** in a bilingual corpus, determining which source language sen- SENTENCE ALIGNMENT tence maps to which target language sentence. This problem has received a lot of attention in recent years and can now be solved with reasonable accuracy (Melamed, 1999) **more alignment citations here** The second problem, word alignment, that is, determining which word(s) of the target correspond to each source language word or phrase, is rather more difficult (Melamed, pear) **more citations here**, and is often addressed with EM methods (cf. Chapter 7). From bilingual corpora aligned in these ways it is possible to count how many times a word, phrase, or structure gets mapped to each of its possible translations. Such alignments are potentially useful not only for MT but also for automatic generation of bilingual dictionary entries for use by human translators (Dagan and Church, 1997; Fung and McKeown, 1997).

Let's now consider an example. Suppose we want to translate the two-word Japanese phrase *2000nen taio* into English. The most probable translation for the first word is, we will assume, *2000*, followed by *year 2000*, *Y2K*, *2000 years*, *2000 year* and some other possibilities. The most probable translation for the second word is, we will assume, *correspondence*, followed by *corresponding*, *equivalent*, *tackle*, *deal with*, *dealing with*, *countermeasures*, *respond*, *response*, *counterpart*, *antithesis* and so on. Thus, according to the translation model alone, the most highly ranked candidate will be the composition of the most highly ranked words, namely *2000 countermeasures*. But, when the contribution of the fluency model, perhaps a bigram model, is factored in, the translation *dealing with Y2K* will have the highest overall score.

Of course, more complex translations models are possible: all that is required is that they generate multiple translations with a ranking associated with each. It is even possible to do "multi-engine" translation, where several translation models (for example a powerful but brittle interlingua-based one and a robust but low-quality direct one) are run in parallel to generate various translations and translation fragments, with the final output determined by assembling the pieces which have highest confidence scores (Brown and Frederking, 1995).

Search

So far we have a theory of which sentence is best, but not of how to find it. The problem is that the number of possible translations are enormous, and so we must find the best output without actually generating the infinite set of all possible translations. But this is just a decoding problem, of the kind we have seen how to solve via the pruned Viterbi (beam-search) or A* algorithm of Chapter 7. For MT this decoding is done in the usual way: outputs (translations) are generated incrementally, and evaluated at each point. If at any point the probability drops below some criterion that line of attack is pruned. Generation can be left to right or outward from heads.

Good introductions to statistical MT include (Brown *et al.*, 1990) and (Knight, 1997). One of the most influential recent systems is described in (Knight *et al.*, 1994).

20.6 USABILITY AND SYSTEM DEVELOPMENT

Since MT systems are generally run by human operators, the human is available to help the machine. One way to use human intervention is interactively; that is, when the system runs into a problem, it can ask the user. For example, a system given the input *the chicken are ready to eat* could, generate paraphrases of each possible meaning, and present the user with those alternatives, for example, asking her to decide whether the sentence means *the chicken are ready to be eaten* or *the chicken are ready to eat something*. It turns out that this is incredibly annoying — users do not like to have to answer questions from a computer, or to feel that they exist to help the computer get its work done (Cooper, 1995). On the other hand, people are very comfortable with the job of fixing up poorly-written sentences, and so post-editing is the normal mode of human interaction with MT systems.

People are also able to edit sentences of the source language, and this ability can be exploited as way to improve the translatability of the input by simplifying it in various ways. Such **pre-editing** can be more cost-effective than post-editing if a single document needs to be translated into to several languages, then the cost of pre-editing is amortized over many output languages — as is often the case for companies which sell things, plus their associated documentation, in many countries (Mitamura and Eric H. Nyberg, 1995). In order to decide what needs pre-editing, one way is to apply MT and see what comes out wrong, and then go back and rewrite those sentences in the original. Another way is to have a model of what MT ought to handle, and require input sentences to be rewritten in that sublanguage, for example, by disallowing PPs which could attach ambiguously. If such a model exists, the pre-editing phase can be dispensed with, by training the technical writers to only write in simple, unambiguous **controlled language**, a version of English, that passes the constraints of the sublanguage grammar checker.

PRE-EDITING

CONTROLLED LANGUAGE

Doing so may also make the source language text more understandable. This is interesting as a case where focusing on the larger task (getting information from tech writers to customers), rather than the problem as originally posed (to translate some existing documents), leads to improvements of the entire process.

In general, user satisfaction is vital for MT systems. Various evaluation metrics are used to predict acceptability. Evaluation metrics for MT intended to be used raw (for information acquisition) include the percentage of sentences translated correctly, or nearly correctly, where correctness depends on both fidelity and fluency. One evaluation metric for MT output to be post-edited is the **edit cost**, either via some automatic measure of edit-distance from a standard translation (such as was introduced in Chapter 7 for evaluating speech recognition), or a direct measure of the amount of time (or number of keystrokes) required to correct the output to an acceptable level.

In general the content words are crucial; users can generally recover from scrambled syntax, but crucially depend on having the words translated. In practice, one of the major advantages of using a MT system is that it handles most of the tedious work of looking up words in bilingual dictionaries.[6] As a result, professional MT users put great value on dictionary size and quality. Such users typically augment the basic system dictionary with the purchase of a domain-specific dictionary designed for the type of translation work they do: medical, electronic, financial, military intelligence etc. But no off-the-shelf dictionary, even one developed from a corpus of texts in the proper domain area, is more than an approximation to the dictionary needed by a specific customer, and so established translation bureaus typically invest substantial effort in augmenting the system dictionaries with entries of their own. The structure of these dictionaries is simple because the specialist **terminology** of any field is generally unambiguous — a photon is a photon is a photon, no matter what context it comes up in — and because terminology is almost invariably open-class words, with no syntactic idiosyncrasies.

TERMINOLOGY

User satisfaction also turns out to depend on factors other than the actual quality of the translation. Many users care less about output quality than other factors, such as cost, speed, storage requirements, the ability to run transparently inside their favorite editor, the ability to preserving SGML tags, and so on. **Translation memory**, the ability to store and recall previously corrected translations, is also a big selling point.

TRANSLATION MEMORY

[6] MT systems can also help in typing in the target language word, especially for translations into Chinese and Japanese, where it is time-consuming to enter characters.

Although for expository purposes the previous sections have focussed on a few basic problems that arise in translation, it is important to realize that these far from exhaust the things that MT systems have to worry about. As Section 20.1 may have suggested, language differences are a virtually inexhaustible source of complexity; and if you were reading the footnotes in the previous sections, you may have been annoyed that every "fact" we mentioned about a language was actually an oversimplification. Indeed, much of the work developing a MT system is down in the weeds, dealing with details like this, regardless of the overall system architecture chosen. Furthermore, adding more knowledge does not always help, since a working MT system, like any huge software system, is a large, delicate piece of code. Improvement to the treatment of one phenomenon, or a correction of a bug in the translation of one sentences, can cause other sentences, previously translated correctly, to go awry.

Given all this, it is surprising that MT systems so as well as they do. One development technique of proven value is iterative development: build it, evaluate it in actual use, improve it, and repeat. In the course of this process the MT system is adapted to a domain, the working habits of its users, and the needs of the consumers of the output.

20.7 Summary

- Although MT systems exploit many standard language-processing techniques, there are also some MT-specific ones, including notably syntactic transformations.
- We have presented four models for MT, the **transfer**, **interlingua**, **direct**, and **statistical** approaches. Practical MT systems today, however, typically combine ideas from several of these models; and MT research systems are probing other niches in the design space.
- MT system design is hard work, requiring careful selection of models and algorithms and combination into a useful system. Today this is more a craft than a science, especially since this must be done while minimizing development cost.
- While MT system design today is thus fairly ad hoc, there are ongoing efforts to develop formal models of the process (Alshawi *et al.*, 1998; Knight and Al-Onaizan, 1998; Wu and Wong, 1998)
- While the possibilities for improvement for MT is truly impressive,

the output of today's systems is acceptable for **rough translations** for information-acquisition purposes, **draft translations** intended to be post-edited by a human translator, and translation for **sublanguage** domains.

- As for many software tasks, user interface issues in MT are crucial; the value of MT systems to users is not directly related to the sophistication of their algorithms or representations, nor even necessarily to output quality.
- Despite half a century of research, MT is far from solved. Human language is a rich and fascinating area whose treasures have only begun to be explored.

BIBLIOGRAPHICAL AND HISTORICAL NOTES

Work on models of the process and goals of translation goes back at least to Saint Jerome in the fourth century (Kelley, 1979). The development of logical languages, free of the imperfections of human languages, for reasoning correctly and for communicating truths and thereby also for translation, has been pursued at least since the 1600s (Hutchins, 1986).

By the late 1940s, scant years after the birth of the electronic computer, the idea of MT was raised seriously (Weaver, 1955) In 1954 the first public demonstration of a MT system prototype (Dostert, 1955) led to great excitement in the press (Hutchins, 1997). The next decade saw a great flowering of ideas, prefiguring most subsequent developments. But this work was ahead of its time — implementations were limited by, for example, the fact that pending the development of disks there was no good way to store dictionary information.

As high quality MT proved elusive, a growing consensus on the need for more basic research in the new fields of formal and computational linguistics (Bar-Hillel, 1960) led in the mid 1960s to a dramatic cut in funding for MT research. As MT research lost academic respectability, the Association for Machine Translation and Computational Linguistics dropped MT from its name. Some MT developers, however, persevered, slowly and steadily improving their systems, and slowly garnering more customers. Systran in particular, developed initially by Peter Toma, has been continuously improved over 40 years. Its earliest uses were for information acquisition, for example by the US Air Force for Russian documents; and in

1976 an English-French edition was adopted by the European Community for creating rough and post-editable translations of various administrative documents. Our translation example in the introduction was produced using the free Babelfish version of Systran on the Web. Another early successful MT system was Météo, which translated weather forecasts from English to French; incidentally, its original implementation (1976), used "Q-systems", which later evolved into Prolog, the unification-based programming language.

The late 1970s saw the birth of another wave of academic interest in MT. One source of excitement was the possibility of using Artificial Intelligence techniques ideas, originally developed for story understanding and knowledge engineering (Carbonell *et al.*, 1981). This interest in meaning-based techniques was also a reaction to the dominance of syntax in computational linguistics at that time. Another motivation for the use of interlingual models was their introspective plausibility: the idea that MT systems should translate as people do (presuming that people translate by using their ability to understand). Introspection here may be misleading, since the process of human translation is enormously complex and furthermore the relevance for machine translation is unclear. Concerns about such issues were much discussed in the late 1980s and early 1990s Tsujii (1986), Nirenburg *et al.* (1992), Ward (1994), Carbonell *et al.* (1992). Meanwhile MT usage was increasing, fueled by the increase in international trade and the growth of governments with policies requiring the translation of all documents into multiple official languages, and enabled by the proliferation of word processors, and then personal computers, and then the World Wide Web.

The 1990s saw the application of statistical methods, enabled by the development of large corpora. Excitement was provided by the "grand challenge" of building speech-to-speech translation systems (Kay *et al.*, 1992), (Frederking *et al.*, pear) where MT catches up with the modern vision of computers being embedded, ubiquitous and interactive. On the practical side, with the growth of the user population, user's needs have had an increasing effect on priorities for MT research and development.

Good survey of the early history of MT are Hutchins (Hutchins, 1986) (Hutchins, 1997). The textbook by Somers and Hutchins (Hutchins and Somers, 1992) includes a wealth of examples of language phenomena that make translation difficult, and extensive descriptions of some historically significant MT systems.

Academic papers on machine translation appear in the journal *Machine*

Translation and in the proceedings of the biennial (odd years) Conferences on Theoretical and Methodological Issue in Machine Translation.

Reports on systems, markets, and user experiences can be found in *MT News International*, the newsletter of the International Association for Machine Translation, which is the umbrella organization for the three regional MT societies: the Association for MT in the Americas, the Asian-pacific Association for MT, and the European Association for MT. These societies have annual meetings which bring together developers and users. The proceedings of the biennial MT Summit (odd years) are also often published. The mainstream computational linguistics journals and conferences also occasionally report work in machine translation.

EXERCISES

20.1 Select at random a paragraph of Chapter 9 which describes a fact about English syntax. a) Describe and illustrate how your favorite foreign language differs in this respect. b) Explain how a MT system could deal with this difference.

20.2 Go to the literature section of the library, and find a foreign language novel in a language you know. Copy down the shortest sentence on the first page. Now look up the rendition of that sentence in an English translation of the novel. a) For both original and translation, draw parse trees. b) For both original and translation, draw dependency structures. c) Draw a case structure representation of the meaning which the original and translation share. d) What does this exercise suggest to you regarding intermediate representations for MT?

20.3 Pick a word from the first sentence of the top article of today's newspaper. a) List the possible equivalents found in a bilingual dictionary. b) Sketch out how a MT system could choose the appropriate translation to use based on the context of occurrence. c) Sketch out how this could be done without using contrastive knowledge.

20.4 The idea of example-based MT can be extended to "translation by analogy" (Sato and Nagao, 1990). a) Given the bilingual data in Figure 20.11, what Japanese word do you think would be appropriate as a translation of *on* in *research on gastropods*? b) Specify an algorithm for doing lexical transfer in this way. Hint: recall the techniques for information retrieval in Chapter 16 ? c) How is your approach similar to choice of TL words by using a TL language model (Section 20.5)? d) How is it similar to disambiguation using semantic features as in Chapter 16?

the cat **on** the mat	no ue no
more notes **on** decision making	ni tsuite
pink frosting **on** the cake	no
see boats **on** the pond	no, ni
always reading **on** the bus	de

Figure 20.11 A mini-corpus of made-up phrases involving *on* and their Japanese translations

20.5 Type a sentence into a MT system (perhaps a free demo on the Web) and see what it outputs. a) List the problems with the translation. b) Rank these problems in order of severity. c) For the two most severe problems, suggest the probable cause of the problem.

20.6 Since natural languages are hard to deal with, due to ambiguities, irregularities, and other complexities, it is much nicer to work with something with is more logical: something that does not have these ‘flaws’ of natural language. As a result, various notations which are (in some ways) less ambiguous or more regular than English have been proposed. In addition to various meaning representation schemes, natural languages such as Esperanto and Sanskrit, have also been proposed for use as interlinguas for machine translation. Is this a good idea? Why or why not?

20.7 Consider the types of ‘understanding’ needed: 1. for a natural language interface to a database, as seen in Chapter 15 2. for an information extraction program, as seen in Chapter 15. 3. for a MT system. Which of these requires a deeper understanding? In what way?

20.8 Choose one of the generation techniques introduced in Chapter 19 and explain why it would or would not be useful for MT.

20.9 Version 1 (for native English speakers): Consider the following sentence:

> These lies are like their father that begets them; gross as a mountain, open, palpable.

Henry IV, Part 1, act 2, scene 2

Translate this sentence into some dialect of modern vernacular English. For example, you might translate it into the style of a New York Times editorial or an Economist opinion piece, or into the style of your favorite television talk-show host.

Version 2 (for native speakers of other languages): Translate the following sentence into your native language.

> One night my friend Tom, who had just moved into a new apartment, saw a cockroach scurrying about in the kitchen.

For either version, now:

a) Describe how you did the translation: What steps did you perform? In what order did you do them? Which steps took the most time? b) Could you write a program that would translate using the same methods that you did? Why or why not? c) What aspects were hardest for you? Would they be hard for a MT system? d) What aspects would be hardest for a MT system? are they hard for people too? e) Which models are best for describing various aspects of your process (direct, transfer, interlingua or statistical)? f) Now compare your translation with those produced by friends or classmates. What is different? Which is better? Why were the translations different?

20.10 Newspaper reports of MT systems invariably include an example of a sentence, typically a proverb, that when translated from English to language X, and then back to English, came out funny. a) Is this evidence that at least one of the two MT systems was bad? b) Why does this problem not arise with human translators? or does it? c) On the other hand, does a successful translation to a foreign language and back indicate that the system is doing well?

20.11 Set yourself an information acquisition task: for example, to find a World-Wide Web page in your favorite foreign language reviewing a recent movie, and discover what the reviewer thought. Accomplish this task using one or two of the Web's machine translation providers. a) Give two examples each of correct and incorrect translations you encountered. b) Come up with a simple quality metric for rating the MT output, and use it to evaluate the MT systems you tried. c) Were you able to find a page of the kind you wanted? d) Were you able to figure out whether the reviewer liked the movie? e) Were the scores on your quality metric predictive of your answers to (c) and (d)?

20.12 Consider each of the following as an application for machine translation. Rank the difficulty of each from 1 (easy) to 4 (very very hard). Also, for each task, say briefly what makes it easy or hard.

a. letters between an American girl and her Chinese pen-pal
b. electronic mail
c. articles in chemistry journals
d. magazine advertisements
e. children's storybooks
f. history books
g. an English-speaker wanting to read articles in Japanese newsgroups
h. an English-speaker wanting to post articles to a Japanese newsgroup

Alshawi, H., Bangalore, S., and Douglas, S. (1998). Automatic acquisition of hierarchical transduction models for machine translation. In *Proceedings of* COLING/ACL-98, pp. 41–47.

Bar-Hillel, Y. (1960). The present status of automatic translation of languages. In *Advances in Computers* **1**, pp. 91–163.

Barnstone, W. (1993). *The Poetics of Translation*. Yale University Press.

Brown, P. F., Cocke, J., Pietra, S. A. D., *et al.* (1990). A statistical approach to machine translation. *Computational Linguistics*, *16*(2), 79–85.

Brown, P. F., Della Piettra, S. A., Della Piettra, V. J., and Mercer, R. L. (1993). The mathematics of statistical machine translation: Parameter estimation. *Computational Linguistics*, *19*(2), 263–311.

Brown, R. and Frederking, R. (1995). Applying statistical english language modeling to symbolic machine translation. In *6th International Conference on Theoretical and Methodological Issues in Machine Translation*.

Brown, R. D. (1996). Example-based machine translation in the pangloss system. In *Coling-96*, pp. 169–174.

Cao, X. (1973). *The Story of the Stone. Also known as the Dream of the Red Chamber*. Penguin Classics, London. First published in Chinese in 1792, translated into English by David Hawkes.

Carbonell, J. G., Cullingford, R. E., and Gershman, A. V. (1981). Steps toward knowledge-based machine translation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *3*(4), 376–392.

Carbonell, J. G., Mitamura, T., and Eric H. Nyberg, r. (1992). The KANT perspective: A critique of pure transfer (and pure interlingua, pure statistics, ...). In *International Conference on Theoretical and Methodological Issues in Machine Translation*.

Chandioux, J. (1976). MÉTÉO: un système opérationnel pour la traduction automatique des bulletins météorologiques destinés au grand public. *Meta*, *21*, 127–133.

Colmerauer, A. and Roussel, P. (1996). The birth of Prolog. In *History of Programming Languages – II*, pp. 331–352. ACM Press/Addison-Wesley, New York.

Comrie, B. (1989). *Language Universals and Linguistic Typology*. Basil Blackwell, Oxford. Second edition.

Cooper, A. (1995). *About Face: The essentials of user interface design*. IDG Books.

Croft, W. (1990). *Typology and Universals*. Cambridge University Press, Cambridge.

Dagan, I. and Church, K. (1997). Termight: Coordininating humans and machines in bilingual terminology acquisition. *Machine Translation*, *12*, 89–107.

Dostert, L. (1955). The georgetown-i.b.m. experiment. In Locke, W. N. and Booth, A. D. (Eds.), *Machine Translation of Languages: Fourteen Essays*, pp. 124–135. MIT Press.

Frederking, R., Rudnicky, A., and Hogan, C. (1999 to appear). Interactive speech translation in the diplomat project. *Machine Translation*.

Fung, P. and McKeown, K. (1997). A technical word- and term-translation aid using noisy parallel corpora. *Machine Translation*, *12*, 53–87.

Hofstadter, D. R. (1997). *Le Ton beau de Marot*. Basic Books, New York.

Hutchins, J. (1997). From first conception to first demonstration: the nascent years of machine translation, 1947–1954. a chronology. *Machine Translation*, *12*, 192–252.

Hutchins, W. J. and Somers, H. L. (1992). *An Introduction to Machine Translation*. Academic Press.

Hutchins, W. J. (1986). *Machine Translation: Past, Present, Future*. Ellis Horwood, Chichester, England.

Kay, M. (1984). Functional unification grammar: A formalism for machine translation. In *Proceedings 10th COLING*, pp. 75–79.

Kay, M. (1997). The proper place of men and machines in language translation. *Machine Translation*, *12*, 3–23. first appeared as a Xerox PARC Working paper in 1980.

Kay, M., Gawron, M., and Norvig, P. (1992). *Verbmobil: A Translation System for Face-to-Face Dialog*. CSLI.

Kelley, L. G. (1979). *The True Interpreter: A History of Translation Theory and Practice in the West*. St. Martin's Press, New York.

Knight, K. (1997). Automating knowledge acquisition for machine translation. *AI Magazine*, *18*(4), 81–96.

Knight, K. and Al-Onaizan, Y. (1998). Translation with finite-state devices. In Farwell, D., Gerber, L., and Hovy, E. (Eds.), *Machine Translation and the Information Soup*, pp. 421–437. Springer.

Knight, K., Chander, I., Haines, M., Hatzivassiloglou, V., Hovy, E., *et al.* (1994). Integrating knowledge bases and statistics in MT. In *Proceedings of the Conference of the Association for Machine Translation in the Americas*, pp. xxx–xxx.

Lakoff, G. (1987). *Women, Fire, and Dangerous Things*. University of Chicago Press, Chicago.

Levin, L., Gates, D., Lavie, A., and Waibel, A. (1998). An interlingua based on domain actions for machine translation of task-oriented dialogues. In *International Conference on Spoken Language Processing*, pp. 1155–1158.

Melamed, I. D. (1999). Bitext maps and alignment via pattern recognition. *Computational Linguistics*, *25*(1), 107–130.

Melamed, I. D. (1999 to appear). Word-to-word models of translational equivalence. *Computational Linguistics*.

Mitamura, T. and Eric H. Nyberg, r. (1995). Controlled english for knowledge-based mt: Experience with the KANT system. In *6th International Conference on Theoretical and Methodological Issues in Machine Translation*.

Nichols, J. (1986). Head-marking and dependent-marking grammar. *Language*, *62*(1), 56–119.

Nirenburg, S., Carbonnell, J., Tomita, M., and Goodman, K. (1992). *Machine Translation: A Knowledge-based Approach*. Morgan Kaufmann.

Oard, D. W. (1997). Alternative approaches for cross-language text retrieval. In *AAAI Spring Symposium on Cross-Language Text and Speech Retrieval*.

Sato, S. and Nagao, M. (1990). Toward memory-based translation. In *Proceedings 13th COLING*, pp. 247–252.

Slobin, D. I. (1996). *Two Ways to Travel*, pp. 195–220. Clarendon Press, Oxford.

Steiner, G. (1975). *After Babel*. Oxford University Press, Oxford.

Sumita, E. and Iida, H. (1991). Experiments and prospects of example-based machine translation. In *ACL*, pp. 185–192.

Talmy, L. (1985). Lexicalization patterns: Semantic structure in lexical forms. In Shopen, T. (Ed.), *Language Typology and Syntactic Description, Volume 3*. Cambridge University Press. also UC Berkeley Cognitive Science Program Report No. 30, 1980.

Talmy, L. (1991). Path to realization: a typology of event conflation. In *Proceedings of the 17th Annual Meeting of the Berkeley Linguistics Society*, pp. 480–519 Berkeley, CA.

Tsujii, J. (1986). Future directions of machine translation. In *Proceedings 10th COLING*, pp. 655–668.

Ward, N. (1994). *A Connectionist Language Generator*. Ablex.

Waugh, L. R. (1976). The semantics and paradigmatics of word order. *Language*, *52*(1), 82–107.

Weaver, W. (1955). Translation. In Locke, W. N. and Booth, A. D. (Eds.), *Machine Translation of Languages: Fourteen Essays*, pp. 15–23. MIT Press, Cambridge. originally circulated in 1949.

Wu, D. and Wong, H. (1998). Machine translation with a stochastic grammatical channel. In *Proceedings of* COLING/ACL-98, pp. 1408–1414.