



ELSEVIER

Language & Communication 24 (2004) 59–75

LANGUAGE
&
COMMUNICATION

www.elsevier.com/locate/langcom

A formulaic approach to translation at the post office: reading the signs

Alison Wray^{a,*}, Stephen Cox^b, Mike Lincoln^b,
Judy Tryggvason^b

^aCentre for Language and Communication Research, Cardiff University, P.O. Box 94, Cardiff CF10 3XB, UK

^bSchool of Computing Sciences, University of East Anglia, Norwich NR4 7TJ, UK

Abstract

TESSA is an interactive translation system designed to support transactions between a post office clerk and a deaf customer. The system translates the clerk's speech into British Sign Language (BSL), displayed on a screen, using a specially-developed avatar (virtual human). TESSA is a context-constrained exemplification of one of two basic approaches to machine translation, neither of which can currently fulfil all of the demands of successful automatic translation. Drawing on recent research in theoretical psycholinguistics, we show how TESSA is a convincing prototype model of one aspect of real human language processing. Ways are suggested of exploiting this parallel, potentially offering new possibilities for the future design of artificial language systems.

© 2003 Elsevier Ltd. All rights reserved.

Keywords: Sign language; Machine translation; Formulaic language; Language processing; Multiword strings

1. Introduction

The Holy Grail of automatic machine translation is a device, preferably fully portable, into which a speaker of language A can talk naturally on any subject, and out of which will come an accurate and idiomatic translation in language B. The path taken towards designing such a system is determined by two basic factors. The first is one's belief about how language works, and hence about what is being modelled. The second is expediency in the face of the practical challenges of systems design: what can be done, to what level of success? In many cases, these two factors

* Corresponding author. Tel.: +44-2920-874243; fax: +44-2920-874242.

E-mail address: wraya@cf.ac.uk (A. Wray).

are irreconcilable, as the desired linguistic system is resistant to modelling, while systems capable of being modelled fall far short of what is required linguistically.

However, the translation tool that we describe here is, we believe, evidence that there is a way forward. It comes about because it reflects a model of language processing that is less demanding in terms of the complexity that it attributes to comprehension and production, and therefore is easier to emulate. While both the system and the theoretical model acknowledge, and indeed accommodate, the full creative potential of human language, that potential is treated as secondary rather than primary. As a result, the high visibility of a small but persistent measure of complexity does not eclipse the overriding tendency for language to be very strikingly predictable in form.

1.1. Units and rules in automatic translation

All translation entails the mapping of units of the input language onto units of the output language, and we may therefore envisage approaches to automatic translation as positioned on a continuum. The success of an automatic translation depends on the quality of the unit inventory and on how the rule system combines those units. Since the internal composition of units is undisturbed, it follows that large units such as whole sentences and phrases, each associated holistically with a specified translation, are guaranteed to feature a high level of accuracy, because their internal composition is predetermined, rather than subject to the operation of potentially imperfect combinatory rules. Combining such units may, certainly, still be subject to shortcomings, but they will be largely at the stylistic level, and thus unlikely to create major problems with the message propositions. However, since natural human language is unpredictable in content, it is generally considered ineffective or uneconomical to allow the computer to operate with very large units. Even in a limited domain, one would, it would seem, need to list, as a separate item, every conceivable utterance, along with its best translation.

The smaller the units, the greater the number of messages that can be processed without specific listing, because the content of phrases and sentences becomes manipulable. But the smaller the units, the more the mode of combination becomes an issue, until, at the single word and morpheme level, an entire set of grammatical rules is required. Only with these rules and a full lexicon of manipulable atomic units, will it be possible for the computer to handle the full range of human linguistic product. Thus, while small units in grammatically legal combinations offer most flexibility in translating unpredicted, and even ungrammatical, input accurately and appropriately, if the rules are not quite right, or if they are unable to capture some features of one or both languages' semantics, there is a danger that the input will be misinterpreted and/or misexpressed, or that the system will simply be unable to translate it.

1.2. An expedient approach to machine translation

For as long as small units, combined by rule, are resistant to modelling, approaches that rely on larger units are fair game. However, standard linguistic theory

inevitably views them as a short-term solution, pending the 'cracking' of the code that is fully creative human language. We shall challenge this assumption later. TESSA, the system described here, is subject to the advantages and constraints already mentioned: by using large units, predictable input can be translated into appropriate output with great accuracy, for there are fewer operations vulnerable to errors. But, of course, the entire procedure is much less flexible in the face of unanticipated variation. It follows that the system is only as good as the input is predictable. We shall argue later that human language behaviour is rather more predictable than has been generally acknowledged. If so, TESSA could represent rather more than just an expedient solution to a translation problem. It could also be a working model of a real psycholinguistic strategy.

2. TESSA: a formulaic approach to practical translation tasks

TESSA (Cox et al., 2002a,b, 2003) is an interactive translation system designed to support transactions between a post office clerk and a deaf customer. The system translates the clerk's speech into British Sign Language (BSL) and displays the signs on a screen, using a specially-developed avatar (virtual human). The system has been developed with the collaboration of the UK Post Office (formerly *Consignia*), and the research has been extended within *ViSiCAST*, a project supported by funding under EU Framework V. *ViSiCAST* researches a wider range of approaches to virtual signing, aimed at providing information access and services to deaf people. Although presently fully focussed on speech to sign translation, the underlying principles are applicable to speech to speech translation too, so TESSA may, in the future, be developed to facilitate two-way conversations between the post office clerk and the foreign tourist.

2.1. The design of TESSA

TESSA operates in accordance with a basic assumption, that there is a limited range of things that the post office counter clerk needs to say to the customer. We may note here that there is an evident shortfall between 'need' and 'want'—that is, human beings, given their full rein of creativity and their desire for social interaction, may *want* to say all sorts of things, and these are not going to be accommodated here. Perhaps it is best to envisage the situation as comparable to that of visiting a post office in a country where one knows little or none of the language. The need to effect a particular transaction, and the sense of relief at doing so successfully, will normally be ample compensation for not having been able to hold a discussion about the weather.

While assumptions should not be too precipitously made regarding the needs and wants of deaf customers in their local post office, the project under whose auspices TESSA has been designed is focussed on ensuring that the system does address a genuine shortfall in provision for deaf people. While it is clearly the case that deaf customers can manage quite well—they have had to do so up to now—nevertheless

TESSA potentially offers access to a range of information that they may previously have had difficulty obtaining. The most fundamental limitation of TESSA at present is that it is unidirectional, from clerk to customer. Research continues into developing a sign recognition device that will make two-way conversations possible. Progress has been made in recognition of a very limited set of signs, but there is still much exploratory work to be done in establishing non-intrusive ways to track, and interpret, physical movements outside the studio environment.

Turning now to the practical operation of TESSA, Fig. 1 demonstrates the operation of the system when a deaf person enters the post office and requests the service. When the trials of TESSA began, there was advance publicity that the system was being installed in certain post offices, so deaf users would enter expecting to be able to use it. Using gestures, the customer indicates to a clerk that they are deaf and wish to use the system. The clerk's activation of the system sets in train a series of stages, described in more detail below, but summarised here in order to provide a clear picture of the events:

1. The clerk speaks into her headset microphone.
2. The automatic speech recogniser selects possible messages from an extensive set of pre-entered targets, and displays those with the highest probability on a screen.
3. The clerk confirms the message intended.
4. The whole message, or the message components, are translated into a set of sign generation instructions.
5. The signed translation is generated and displayed on a video screen.

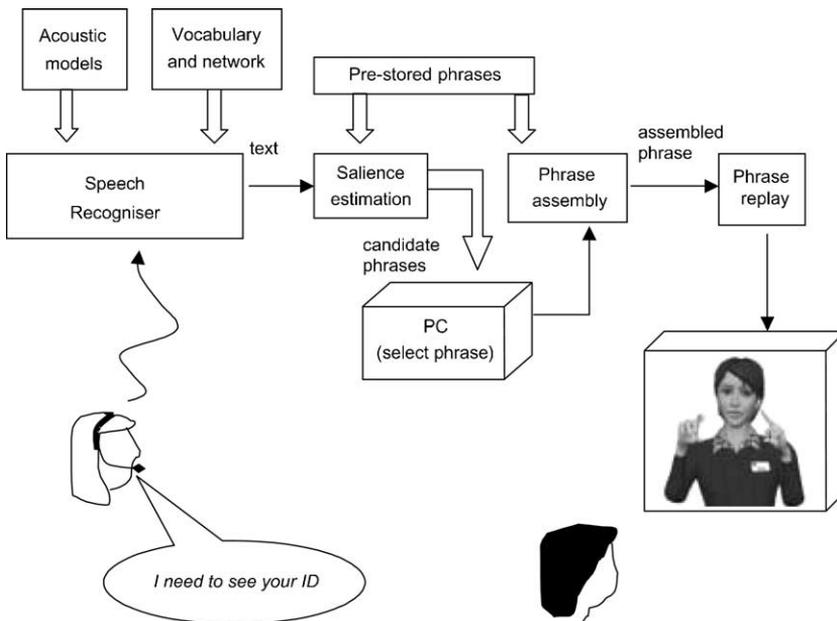


Fig. 1. The Post Office translation system.

Thus, the thrust of the TESSA system is the amalgamation of three major components, each of which is a computing challenge in its own right: speech recognition, translation, and sign generation.

2.2. Speech recognition and matching to the database

The post office clerk, responding appropriately to the transaction initiated by the customer (e.g. passing her a completed form, package or whatever), activates the recogniser by pressing a key, and speaks into the headset microphone. The speech recogniser¹ performs acoustic matching to the clerk's utterance, identifying the most probable items from amongst those stored in its lexicon. The lexicon is derived from recordings of transactions done in real post offices, and from simulation experiments (see below). The items that lexicon contains comprise some single words but also a great many phrases (e.g. *Can you give me; I need to see*)². Each item has associated with it a score for its acoustic components that gives, essentially, the salience of that acoustic sequence for identifying the item. These 'salience' scores are summed for each output phrase, and the system displays the topmost highly scoring phrases to the clerk for confirmation. The speech recogniser is able to handle both natural acoustic variation within the precise wording of a target item, and also a wide range of paraphrases of the target. To illustrate the latter, Table 1 presents alternative formulations of target phrases that were observed during the simulations of post office transactions.

The left-most column in Table 1 indicates the target phrase in the TESSA system, and the other columns represent the kinds of variation that naturally occurred when speakers aimed to produce that message. In the simulations, a subject was asked to imagine that he or she was a post office clerk and to make a request or supply information to a customer. The challenge for TESSA is to ensure that any of these paraphrases will be identified as equivalent to the target. As this is done on an acoustic basis only—there is no dictionary or thesaurus—success is dependent on common properties between the target and the paraphrase. Thus, the most successfully recognized paraphrases of *Can you complete another deposit slip please?* are likely to be those that feature acoustic sequences consistent with *complete*, *deposit* and *slip*. The probability calculations help to ensure that material that rarely occurs elsewhere is counted as more salient than material that occurs in many phrases. As a result, the system will prioritise the occurrence of *slip* over *please* and *you*.

The five target phrases with the highest accumulated salience are displayed to the operator, who selects the best match, or, if none of them is correct or close enough, the operator can, at that stage, re-word her input to the speech recogniser. With this

¹ The speech recogniser will not be described here, but details can be found in Cox et al. (2002a).

² In order to maximize accuracy, the recogniser contains acoustic entries for *both* the multiword strings used for the translation *and* their component words, along with common collocations for some of the most common words. This helps the recogniser to correctly identify input, given its naïve, agrammatical approach (that is, it potentially tolerates any sequence of its lexical items). It will favour interpretations that are attested within multiword strings and also as coded collocates, over word sequences that are never attested.

Table 1
Alternative expressions of phrases used in TESSA

TESSA Sample	Paraphrase 1	Paraphrase 2	Paraphrase 3	Paraphrase 4	Paraphrase 5
Where is it going to?	Where are you sending this?	Where will it go to?	Where is the letter going?	Where is the destination?	Where is the item going?
Can you complete another deposit slip please?	You must fill in another deposit slip.	Could you please complete another deposit slip?	Could you complete another deposit slip please?	Could you fill in this deposit slip?	Can you fill another slip in?
Goodbye.	See you later.	Cheerio.	Goodbye.	Thank you, bye bye.	Bye.
Is there anything else?	Do you require anything else?	Is that it?	Can I help you with anything else?	Can I help you with anything else?	Will that be all?
A charge is made for this service.	There is a charge for this service.	You'll be charged for the service.	You will be charged for this service.	You will be charged for this service.	You will have to pay for this service

checking procedure, there is much less chance that the incorrect message will be translated. It also provides a useful alternative input point. Should it be useful to bypass the speech recogniser, it is possible to call up lists of phrases according to headings such as ‘Postage’, ‘DVLA’,³ ‘Bill Payments’, ‘Passports’, and to select directly from the screen the phrase for translation.

2.3. Translation

The principle upon which TESSA operates is semantically-based translation, in which the primary focus is the meaning of the whole message. Other systems using this approach are the JANUS and DIPLOMAT projects based at Carnegie Mellon University (e.g. Waibel, 1996; Silberman, 2000) and an experimental multilingual speech translation system under development in Japan (e.g. Matsui et al., 2001). The philosophy under which these systems operate is that

A successful speech translation system...cannot rely on perfect recognition or perfect syntax. Rather, it must search for a semantically plausible interpretation of the speaker’s intent while judiciously ignoring linguistically unimportant words or fragments (Waibel, 1996:42).

Where a full parser would trip up in the attempt even to ascertain which words or fragments were ‘linguistically unimportant’, semantically brokered message processing takes no account of the form other than as a mechanism for accessing meaning.

There are many different ways of expressing any given message, but the probability approach, applied to the utterance as a whole, is an effective way of gathering them together as versions of the same basic proposition. Most importantly, because the interpretation is based on making the best match, unanticipated variants and speech errors can be accommodated, provided sufficient salient content is identifiable. Of course, because no direct account is being taken of word order, morphology or other grammatical signals, errors of interpretation can occur. However, they are at least similar to those that would occur in natural human translation if, say, the environment was too noisy for everything to be heard. The potential for misunderstanding is reduced by the on-screen feedback loop to the speaker, who confirms that the correct interpretation has been made.

The TESSA system currently contains a corpus of c.500 target phrases, drawn up by analysing transcripts of post office business (Cox, 2002; Cox et al., 2002b). Each of these phrases has its own allocated assembly and replay instructions, which create a translation in BSL.⁴ TESSA is designed to by-pass complexity at semantic/structural

³ Driver and Vehicle Licensing Agency.

⁴ BSL is used in TESSA even though a less technically demanding solution would be to display a version of the clerk’s speech in written English. For many people who have been profoundly deaf from an early age, signing is their first language, and English (which they use only in the written medium) is their second. As a result, many deaf people have below-average reading abilities for English text and an English text display would be a poor substitute for signs in their native BSL. It is a popular misconception that sign languages are just gestured versions of spoken languages. Rather, British Sign Language is a fully

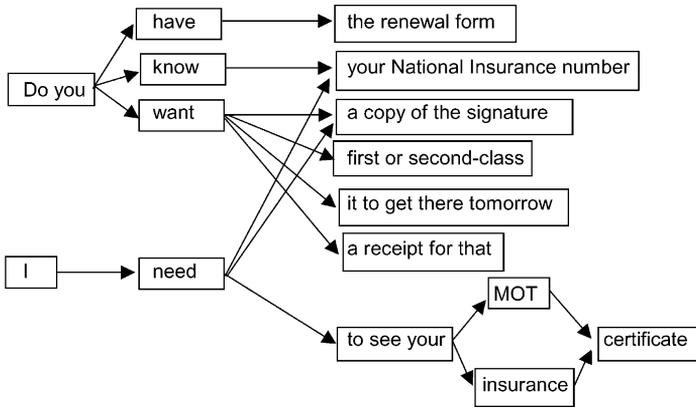


Fig. 2. A section of the recognition network in TESSA.

level, and it does not process individual morphemes or words. It has a grammar, but it is very simple. There are basically two operations, concatenation (Fig. 2), and single level insertion, the latter used to accommodate numerical and monetary amounts: the appropriate numbers, as identified in the input, are slotted into a fixed frame, such as *That's [amount] please*.

In short, TESSA operates essentially as a data look-up system, and is constrained to treat all input as an attempt at one of its target phrases. Whatever the actual form of the input message, once it is recognised as a version of the target message, it will be mapped onto that target and realised as the same form in British Sign Language. That is, all five example paraphrases of a phrase in Table 1 would be translated identically, with no recognition of any subtle differences in emphasis, pragmatics or idiolect that might account for the speaker's choice of one formulation over another.

2.4. Output sign generation

The on-line output assembly entails the use of an avatar (virtual human) that signs the required chunks of message based on the 'motion-captured' movements of human models. The human informants had sensors fixed to their bodies, especially the hands and face, to enable the mapping of the precise movements entailed in making a signed sequence. These movements, stored as representations of a phrase or word, are retrieved and assembled on-line to create the required translation. Although it may seem easier simply to store a video clip of an informant signing the phrases, there are several reasons for not doing this. Firstly, less memory is required in our method than would be required for video clips, since the bandwidth of the

developed language in its own right, largely independent of English, with its own signs to express distinct concepts and with its own syntactic and semantic structures, including prosodic features, and parallel morphological components, that cannot be adequately handled in a fully segmental approach (e.g. use of signing space, position of eyebrows, etc.) (Cox et al., 2002b). This being the case, it is inappropriate simply to translate each English word separately into a sign, and the procedures required of TESSA are in every way equivalent to those required for translating English into a foreign language.

signal for instructing the avatar is much smaller. Secondly, as Fig. 2 indicates, many translations are composed of more than one phrase. The avatar can concatenate message units smoothly, whereas the precise position of the signer across video clips may not align, making the composite message more difficult to follow. Thirdly, it becomes possible to use different signer informants to create the motion-captured lexicon of signs, without disturbing the illusion that a single character is signing. Fourthly, on-line message creation is transferable to future systems. The ultimate aim is to handle unconstrained content, at which point video clips would be unviable. The current approach is sufficiently flexible to operate with a full analytic language processor (see later).

3. Constrained flexibility: the psycholinguistic parallel

There are lots of things that TESSA simply cannot do. It is a highly constrained system, operating only in a specific environment, and with very modest goals. However, it has some strengths, too, both in terms of what it achieves and, as we shall now demonstrate, in reflecting a possible linguistic processing strategy in humans. In the sections below, we first briefly describe the psycholinguistic model with which we draw our parallels. We then demonstrate the points of similarity explicitly. Finally, we use the model to explore where the future might lie in the design of automatic translation systems.

3.1. *Wray's dual systems processing model*

Although traditional linguistic theory adequately demonstrates that human language can be largely encoded and decoded by applying grammatical rules to a lexicon of words and morphemes, there has been, for many years, a recognition that describing language in these terms alone is inadequate. Wray (2002a, chapter 1) quotes writings dating back as far as the mid-nineteenth century and encompassing the works of Saussure, Jespersen, Bloomfield and Firth, to the effect that 'our language does not expect us to build everything starting with lumber, nails, and blueprint, but provides us with an incredibly large number of prefabs' (Bolinger, 1976:1). Such views extend well beyond the minimal recognition, also made in Chomskyan syntax, that non-compositional idioms must be stored holistically in the lexicon. Rather, they also allow for the holistic storage and retrieval of wordstrings of regular construction. Since holistic storage certainly cannot be seen as a replacement for componential processing, it can only, at most, be an alternative to it. Any regular holistically processed string could also be built up from, and broken down into, its parts. As Saussure, Jespersen and Bolinger all suggest, the incentive for not taking this latter approach is probably the saving of processing effort.

Wray (1992, 1998, 1999, 2000, 2002a, b, c, Wray & Perkins, 2000) proposes that humans hold in balance, on the one hand, a creative, analytic approach to language processing, and, on the other, a 'short-cut' system that uses prefabricated, holistically-handled, units, often several words long, to fast-track both production and

comprehension in predictable situations. The intricate universal and local rules of the analytic approach ensure that any message can be created or understood. However, by seeing that approach to processing as one of two, it is possible, without denying that humans *possess* the ability to use language creatively, to allow for the possibility that there are both psychological and sociological reasons why they may refrain from employing that creativity most of the time.

Although we shall focus on Wray's formulation, she is not the only one to have proposed such a balance, and she herself acknowledges a great many versions of the same idea (Wray, 2002a: chapter 1). To take just one example, Sinclair (1991: 114), working in the domain of corpus linguistics, has identified two principles in language processing. The first to be applied is the 'idiom principle'. In relation to input, the principle imposes the assumption that two or more words occurring together belong together as a unit, rather than operating independently. Only if that principle is unsuccessful in rendering a meaning does the other, 'open choice', principle apply, enabling the individual to take a closer look at the words and rules. While the open choice principle greatly increases the likelihood of an outcome for unusual material, its use not only compromises processing speed, but also may undermine comprehension at the pragmatic level, for it will tend to render a literal, word-by-word, reading of the target wordstring, which may not always be what was intended.

The explanatory power of a dual systems processing model is considerable, because it combines the advantages of the two systems. Analytic processing can accommodate the way in which linguistic competence enables us to deal with novel material, something not easily explained in terms of a holistic model since pre-fabricated units offer only a restricted range of forms and meanings. On the other hand, holistic processing more easily explains the idiomaticity associated with a preference for some grammatical strings over others; analytic processing, working on its own, would over-generate, identifying as acceptable strings that are not idiomatic (Pawley & Syder, 1983). A model in which the two modes of processing both operate thus provides a means of explaining both what we *can* do with language and what we customarily *do* do with it.

The two systems work closely together, with the holistic system applying by default. Where problems arise with holistic processing, the analytic option cuts in, resolves the problem and cuts out again. In output production, the lexicon is searched for the largest units available that are associated with the intended meaning. In this way, the assembly procedure is kept as simple as possible. In comprehension, input will be analysed only to the extent necessary to release its meaning. If a multiword string is stored in the lexicon with a particular meaning, this is the meaning that will be allocated to it by default. If the meaning first rendered is inappropriate, then further analysis will be done, until a satisfactory solution is found.

Analysis will also be invoked to handle novelty within a predictable frame, but this analysis will be restricted to the fragments that need it. The most common case is where the stored wordstring contains slots for variation. For instance, the wordstring *Mr McGregor is pulling Herbert's leg* is, technically speaking, novel, and we would not expect that it can be decoded entirely holistically. However, as it is constructed around an idiom, and idioms do not successfully succumb to full analysis

(since they are then understood literally), some measure of holistic processing must take place. The solution is to recognise that the idiom is represented by a frame of fixed words with some slots for specified variation: NP_i pull + TENSE NP_j + POSS leg. In such a case, while the frame itself is handled holistically, analytic processing will attend to the novel items and their juxtaposition within the discourse context. Kuiper (1996) describes how, in the context of livestock auctions, the prefabricated frames are spoken so fast as to be all but incomprehensible. Yet, because they are so familiar to both auctioneer and bidders, they successfully serve as pointers to the novel material (age, weight, provenance and price of the animals) inserted into the gaps.

In short, the principle of dual systems processing is that the very powerful analytic processing strategy—which assembles complex units from, and disassembles them into, smaller, ultimately atomic, components by rule—is used only when holistic processing reaches the end of its capability. This balance of operations maximises the opportunities for processing shortcuts, without losing the flexibility to handle unexpected material. The two systems constantly interact in the interests of maximum accuracy at minimum expense.

3.2. Similarities between TESSA and Wray's model

As should now be clear, the database matching procedure in TESSA resembles in some important ways the holistic system in Wray's model. We shall consider the similarities in turn. First, of course, there is the very holistic storage of complete strings, bypassing the need for a complete semantic and grammatical decoding. In the computer system, as in Wray's model, the advantage of matching large, internally complex items rather than their components is that it reduces the number of operations needed to proceed from input to, in the case of TESSA, translated output, or, in the case of the human speaker, comprehension. The benefits of this are twofold: less likelihood of error, since there are fewer procedures to go wrong, and faster processing, since there is less to look up and there are fewer rules to apply. On the other hand, there is only limited scope for accommodating flexibility, considerably restricting the capability of the system to handle novel input or subtle expression in output.

However, there is *some* flexibility. In both Wray's model and TESSA there are holistically managed frames with gaps in, able to take variable material. In this way, even a quite small inventory of fixed strings can be turned into a sizeable set of possible interpretations and outcomes. This halfway house between fixedness and full compositionality renders high dividends in both domains. Many fixed frames admit a very wide range of possible completions—in human language there are constructions like *no sooner X than Y* and *I have known X for N years*), while even in the narrow world of TESSA the frame *that's AMOUNT please* has as many completions as there are purchase costs to be paid in the post office. In neither case would it be parsimonious to store every possible version of the completion independently, yet not to store any of the material holistically would be to impose on the system a much greater on-line processing burden than is necessary.

A further aspect of flexibility is seen in TESSA's probability matching, which, while not accommodating any interpretation and expression of the variation, does at least tolerate it, by disregarding nonsalient differences between actual input and target phrases. Does this, too, have an equivalent in human language processing? Research into the variation found in ostensibly 'fixed' strings (e.g. Tucker, 1996) does suggest that wordstrings may be identified in input on the basis of probability in relation to the occurrence of their most salient components, just as in TESSA. In other words, the individual learns from experience where to expect variation, and, through a highly conservative approach to extrapolation, develops a sense of the range that that variation can take (Wray, 2002a). That is, the observation that a particular slot is open for variation need not lead to the assumption that the entire range of variation predicted by the language's grammar and lexicon will actually occur, so a weighted probability assessment could be entailed in anticipating what will occur in the gap. A conservative approach ensures that observation remains the main arbiter of what is permissible. In this way, we can account for why, to use Tucker's (1996) example, the slot in *I haven't the SUPERLATIVE idea* is not, in fact, equally likely to attract all superlatives, even if the meaning is appropriate and easy to understand. While *faintest*, *slightest* and *foggiest* slip easily into the gap, *smallest*, *tiniest* and *vaguest* seem a little less comfortable, and *weakest*, *roughest* and *feeblest* seem highly unlikely.

In line with this conservative approach, it may be noted that the possible configurations depicted for TESSA in Fig. 2 are restricted, relative to what might be deemed possible in a fully generative system. Although *I need* can link with both *a copy of your signature* and *your National Insurance number*, their equivalence in that context does not generalise, so *Do you know* links only with the second, not the first. Even strings that would make full sense to us, such as *Do you have a receipt for that* are outlawed, because that message is not one that is needed within the confines of the user context as it is defined.

3.3. Differences between TESSA and Wray's model

The key difference between the theoretical models of Wray and Sinclair and the design of TESSA is that the former supplement holistic processing with analytic processing, enabling the speedy encoding and decoding of predictable forms, but the promise of a fuller analysis if necessary. TESSA, of course, does not. In humans the holistic-only system would be too restrictive—consider the problems of trying to survive in a foreign country with only the exact entries in a phrasebook. In the same way, TESSA's holistic look-up system is too restrictive to deal adequately with the many additional messages that might need to be expressed, from time to time, in post office transactions.

What is needed, then, to make up the shortfall in a TESSA-like design, is an analytic processing system. We shall propose, below, that the mere presence of the holistic system considerably alleviates some of the most pressing problems for the design of an analytic processing component. In order to contextualise that proposal, however, we must first briefly review the standard approach to the design of analytic grammars for the computer.

3.4. *The challenges of designing an analytic system*

For many years, systems designers have been challenged to construct a system that emulates the patterns of the human lexicon and grammar. In the context of translation, the task is more than doubled, because both languages need to be mapped separately, but also successfully linked at the semantic level, so as to ensure the accurate transmission of ideas. Key features, then, of a system that aims to offer full compositional language processing are:

1. a complete lexicon of all the morphemes and words that might possibly occur in the input or be needed for output;
2. a set of rules able to encode and decode the words and morphemes into and out of longer, grammatical strings, and to correctly interpret the underlying grammatical form of an ungrammatical realisation in input; and,
3. if necessary, a set of filters to exclude formulations or interpretations that are generated by the grammar but are not desirable.

Despite many years of research, the field of linguistics has still failed to produce a fully satisfactory account of such a language grammar, so it has been difficult to model it. Moreover, there has been increasing objection to the fact that (a) the endeavour excludes performance features and pragmatic and contextual interpretation, and (b) it is fundamentally uncomfortable with fuzzy judgements, collocational restrictions and irregularity. The persistent adoption of the syntactic approach in machine-based language simulation, including translation, reveals a belief that, despite these problems, language can ultimately be characterised in this way, and that doing so will open the road to the comprehension and production of human language by machines (e.g. Taylor et al., 2000).

It may seem at first sight incontrovertible that accessing meaning ultimately depends on unpacking structure. Change a verb tense, a lexical item, the word order, or the structural relationship between two phrases or clauses, and you change the meaning in largely predictable ways. Syntactic research strives to identify a productive rule system that will enable the creation of all, and only, the combinations of words that are grammatical in a language. It has long been believed in many quarters that if machines are to emulate human language processing then they must possess a version of that rule system.

Difficulties have always arisen, however, in pinning down rules that map out accurately the curiosities of language at different levels. In the first place, it is a painstaking business to encode programs adequately at all. ViSiCAST's own generation system (Cox et al., 2002b, Marshall and Sáfár, 2001; Sáfár and Marshall, 2001) clearly illustrates the investment entailed. It is essentially a syntactically driven parser, based upon head-driven phrase structure grammar (HPSG) (Pollard and Sag, 1994). The current prototype has a vocabulary of 50 tokens and a few grammatical rules. Although this modest base has tremendous long-term potential, and even now offers sufficient variation to observe the characteristics of the decoding and encoding that programming brings about, it is

certainly not yet able to match the even limited effectiveness of the phrase-based TESSA system.

One not insubstantial problem is that human languages all display a deal of complexity that might seem surplus to requirements: variation in the patterns of conjugation and case agreement, sub-rules, and irregularities. These make it difficult simply to write a single rule that applies to all potential members of a word-class or larger composite structure.

Meanwhile, the semantic relationships between words and concepts are messy, and languages do not always provide one-to-one associations. Non-alignments between words in different languages can render translation difficult. The source language may have a particular word that conjures up a great many cultural associations that cannot be conveyed adequately by any single word in the target language. For instance, an English speaker will make very different cultural associations if told, on the one hand, *The garden was full of doves* or, on the other, *The garden was full of pigeons*. Yet in German the word *Tauben* is used for both. On the other hand, the English word *you* semantically divides in T-V languages like French and German, so that translating *you* into German requires a decision about whether to use *Du* [familiar (T) singular], *Ihr* [familiar (T) plural] or *Sie* [polite (V) single and plural]. The translator must interpret the neutral *you* according to contextual indications about the nature of the relationship between the person using the pronoun and the person(s) being referred to. Even translations between T-V languages may not assume that the T form in one necessarily calls up the T form in the other, since cultural differences can draw the line between T and V usage in a different place.

One solution to the problem of assigning adequate semantic readings to individual words without making them untranslatable is to create a conceptual hierarchy. Wierzbicka (e.g. 1996) proposes a set of ‘conceptual primitives’ shared by all humans. These form a base level of meaning that can be considered universal. Additional, increasingly language-specific, meaning is added in subsequent layers. Such an approach ensures that there is always a common denominator between languages, making translation possible at whichever layer first unites the concepts. There are, however, potential objections to this approach. One relates to the assumed psychological reality of the primitive concepts. In the case of *pigeons*, *doves* and *Tauben*, the common conceptual layer required for translation would have to be that at which pigeons and doves are recognised to be members of the same genus. However, it is questionable whether speakers of English are at all aware that these two types of bird are the same or, more pertinently, whether their sameness is more fundamental than their differentness.

Whether one favours a direct and ‘flat’ transformation of a linguistic expression into a complex semantic representation, or envisages semantic representation as hierarchical, with primitive concepts at the core, and complexity a function of progressive augmentations of localisation within a particular language, the underlying procedure of drawing meaning from input language form and re-expressing it in output language form remains the same. The combinations of words and morphemes must be sufficiently understood to extract from them the meaning of the

message they convey. Small units and words provide a powerful and flexible tool for this purpose.

Yet, as we have seen, at present, such models fall far short of their potential. Analysts dedicated to modelling language in this way believe that the accuracy of input processing and the naturalness of the output will ultimately improve. In the meantime, the only way to ensure satisfactory target language output is to operate in an environment in which a human can intervene to volunteer information and correct automatic processing.

3.5. Discarding the either/or approach

Our contention is that the approach to analytical grammars described above is based on a false premise: that the grammar has to generate *everything*. We propose that the future for automatic machine translation, as for computer-mediated language processing as a whole, may be the combination of holistic and analytic processing, using the former to reduce the amount of processing necessary, and the latter to make up the shortfall.

But there is much more to our position than just the suggestion that a look-up system might enable the by-passing of some operations that could otherwise be done entirely adequately the long way round by the analytic route. Indeed, the opportunity for short-cutting in this way has long been operational in computer-based systems of many kinds. Rather, we propose that our approach delivers specific benefits in the very design of the analytic component.

The major challenge for an analytic system is borderline irregularity. Minor inconsistencies in a form or meaning relative to the general rule include such things as:

- the absence from English of the words **couth* and **ept* even though they appear to have negative forms in *uncouth* and *inept*;
- partial patterns across morphological forms such as Chomsky's (1965: 186) example *horror—horrid—horrify*; *terror—*terrid—terrify*; *candor—candid—*candify*; and
- any number of unpredictable meanings for pairings of apparently transparent words, e.g. *perfect stranger*; *first thing*; *sleep tight*; *broad daylight*.

Because they place awkward restrictions on the general rule, and these exceptions must be accommodated either in the rule itself or in a *post hoc* filter, examples of this kind place a disproportionate burden on the grammar.

However, Wray's work strongly suggests that a sizeable proportion of such irregularities—at the morphological, lexical and grammatical levels—are actually contained within fixed expressions, that is, within wordstrings that are not generally subjected to analytic processing. Indeed, those oddities are, she suggests, actually a symptom, even a product, of holistic processing (Wray, 2002a: 267). A regular wordstring of, say, the 18th century, passing into the holistic lexicon for convenience, and inherited as a prefabricated phrase from generation to generation,

would be so protected from the kind of regular examination, rationalisation and updating that would occur if grammatical rules were regularly applied, that its retained historical structure or lexical content would become a fossil within the changing language.

It follows that excluding these expressions from the scope of the full analysis will make it much easier to write the rules of the analytic grammar. To a small extent this is already accepted—no grammar-writer would relish creating a special rule to construct the phrase *by and large*, since no other phrase requires the (apparent) coordination of a proposition and an adjective, and there is no way of predicting or working out the meaning from the parts.

But this approach could extend much, much further. Subject to the establishment of reliable criteria for identifying wordstrings as formulaic (see Wray and Namba, 2003, for one attempt) there may, in due course, be available to the research community lists of wordstrings that, despite their apparent innocuousness, should, and can, be exempted from the active rules for grammatical and semantic encoding and decoding. Including in the lexicon of the analytic grammar a large dictionary of prefabricated material, some of it in the form of a frame with gaps, would thus both simplify the grammar itself and provide substantial savings in processing. In both regards it would mirror what is hypothesised to operate in humans.

In due course, there may be further benefits yet. Wray's model of how the child acquires the analytic grammar in the first place might itself be open to exploration in the computational domain, even, perhaps, leading to the development of an input-driven extrapolated grammar and lexicon, in stark contrast to the current practice of rule-writing and a pre-entered lexicon. In fact, the bootstrapping approach is already successful in allied domains, and the plausibility of the emergent system as a genuine model of how humans develop syntax is now presenting a major challenge to Chomsky's proposal that syntactic complexity in human language is due to the chance design parameters of a one-off specialised linguistic module. Computational experiments and evidence from typological studies of languages (see, for instance, studies in Wray, 2002d and Cangelosi and Parisi, 2002) now suggest that, rather, the grammar of human language may be a product of general psychological processes that extrapolate from input data according to simple pattern recognition. Insofar as an artificial emergent system can be guided by the same simple observational principles that it is proposed lead to the acquisition of syntax in humans, it should be possible to develop input-to-output artificial language processors that are, like humans, sufficiently flexible to accommodate novelty, without the generative tail wagging the expediency-driven dog.

References

- Bolinger, D., 1976. Meaning and memory. *Forum Linguisticum* 1, 1–14.
- Cangelosi, A., Parisi, D. (Eds.), 2002. *Simulating the Evolution of Language*. Springer, London.
- Chomsky, N., 1965. *Aspects of the theory of Syntax*. Cambridge, MA: MIT Press.
- Cox, S., 2002. Speech and language processing for a constrained speech translation system. *Proceedings of the International Conference on Spoken Language Processing*, Denver, September 2002.

- Cox, S., Lincoln, M., Tryggvason, J., Nakisa, M., Wells, M., Tutt, M., Abbott, S., 2002a. TESSA, a system to aid communication with deaf people. *ASSETS 2002: Proceedings of the Fifth International ACM SIGCAPH Conference on Assistive Technologies*, Edinburgh, July 2002, pp. 205–212.
- Cox, S., Lincoln, M., Nakisa, M., Wells, M., Tutt, M., Abbott, S., 2003. The development and evaluation of a speech to sign translation system to assist transactions. *International Journal of Human Computer Interaction* 16 (2), 141–161.
- Cox, S., Marshall, I., Sáfár, É., 2002b. What are the difficulties of translating English to Sign Language? *The Linguist* 41 (1), 6–10.
- Kuiper, K., 1996. *Smooth Talkers: the Linguistic Performance of Auctioneers and Sportscasters*. Lawrence Erlbaum Associates, Mahwah, NJ.
- Marshall, I., Sáfár, É., 2001. Extraction of semantic representations from syntactic CMU link grammar linkages. In: Angelova, G., Bontcheva, K., Mitkov, R., Nikolov, Nicolas., Nikolov, Nikolai. (Eds.), *Recent Advantages in Natural Language Processing (RANLP)*. Tzigov Chark, Bulgaria, pp. 154–159.
- Matsui, K., Wakita, Y., Konuma, T., Mizutani, K., Endo, M., Murata, M., 2001. An experimental multilingual speech translation system. *Perceptive User Interfaces Workshop*, Orlando Florida, 15–16 Nov 2001. Available from: <http://www.cs.ucsb.edu/PUI/PUIWorkshop/PUI-2001/a11.pdf>.
- Pawley, A., Syder, F.H., 1983. Two puzzles for linguistic theory: nativelike selection and nativelike fluency. In: Richards, J.C., Schmidt, R.W. (Eds.), *Language and Communication*. Longman, New York, pp. 191–226.
- Pollard, C., Sag, I.A., 1994. *Head-driven phrase structure grammar*. University of Chicago Press, Chicago.
- Sáfár, É., Marshall, I., 2001. Translation of English text to a DRS-based Sign Language oriented semantic representation. In: *Conference sur le Traitement Automatique des Langues Naturelles (TARN)*, vol 2, pp. 297–306.
- Silberman, S., 2000. Talking to strangers. *Wired* 8.05. Available: <http://www.wired.com/wired/archive/8.05/translation.html>.
- Sinclair, J.McH., 1991. *Corpus, Concordance, Collocation*. Oxford University Press, Oxford.
- Taylor, J.G., Taylor, N.R., Apolloni, B., Orovas, C., 2000. Constructing symbols as manipulable structures by recurrent networks. In: *IEEE-INNS-ENNS International Joint Conference on Neural Networks (IJCNN 2000)* vol 2, pp. 2099–2104.
- Tucker, G., 1996. So grammarians haven't the faintest idea: reconciling lexis-oriented and grammar-oriented approaches to language. In: Hasan, R., Cloran, C., Butt, K. (Eds.), *Functional Descriptions: Theory in Practice*. John Benjamins, Amsterdam, pp. 145–178.
- Waibel, A., 1996. Interactive translation of conversational speech. *Computer* 29 (7), 41–48.
- Wierzbicka, A., 1996. *Semantics: Primes and Universals*. Oxford University Press, Oxford.
- Wray, A., 1992. *The Focusing Hypothesis: the Theory of Left Hemisphere Lateralised Language Re-examined*. John Benjamins, Amsterdam.
- Wray, A., 1998. Protolanguage as a holistic system for social interaction. *Language and Communication* 18, 47–67.
- Wray, A., 1999. Formulaic language in learners and native speakers. *Language Teaching* 32 (1), 213–231.
- Wray, A., 2000. Holistic utterances in protolanguage: the link from primates to humans. In: Knight, C., Studdert-Kennedy, M., Hurford, J. (Eds.), *The Evolutionary Emergence of Language*. Cambridge University Press, Stanford, CA, pp. 285–302.
- Wray, A., 2002a. *Formulaic Language and the Lexicon*. Cambridge University Press, Cambridge.
- Wray, A., 2002b. Dual processing in protolanguage: performance without competence. In: Wray, A. (Ed.), *The Transition to Language*. Oxford University Press, Oxford, pp. 113–137.
- Wray, A., 2002c. Formulaic language in computer-supported communication: theory meets reality. *Language Awareness* 11 (2), 114–131.
- Wray, A., 2002d. *The Transition to Language*. Oxford University Press, Oxford.
- Wray, A., Namba, K., 2003. Formulaic language in a Japanese-English bilingual child: a practical approach to data analysis. *Japanese Journal of Multilingualism and Multiculturalism* 9.
- Wray, A., Perkins, M.R., 2000. The functions of formulaic language: an integrated model. *Language and Communication* 20 (1), 1–28.