So what is so interesting about simultaneous interpreting?

Robin Setton

Ecole de traduction et d'interprétation (ETI), Université de Genève

The question of why simultaneous interpreting merits our interest needs to be asked *a priori* because after sixty years of providing a vital daily service to the international community, the activity remains an arcane field of study. This status of the discipline is probably due in equal parts to the occult, not-quite-respectable odour of translation generally, and to the extreme difficulty of capturing SI for research.

Translation is often regarded, particularly by monolingual speakers of dominant languages, as an irritating necessary evil. When sacred texts have had to be translated to evangelical ends, the translations have never quite acquired the status of holy writ; translations of legal documents gain equal status with the original when multiple official languages are constitutionally recognised, but then they are explicitly not to be considered translations. In the folk view of language as encoding thoughts, or as inextricably bound to them, a translation can never be more than an approximation, a pale substitute for the original. Public figures often seem impatient or resentful of the need to go through interpretation. The prospect of imposing an international language is seen by many now as an historic opportunity far outweighing the sad loss of all the other languages of humanity. Philosophical and linguistic inquiry into translation has contributed to this view in concluding that translation can never be perfect (Quine 1960), but only good enough for practical purposes (Keenan 1978). If the mainstream linguistic and cognitive sciences have not looked to translation or interpretation as valid objects of study, it is probably due to a belief that we can learn about language and thought only by studying the 'pure' products of spontaneous sovereign linguistic creation from spontaneous thought.

The analyses and conclusions of authors like Quine and Keenan are valid, but the mistaken inference that translations are therefore products of linguistic communication which are less pure than spontaneous production, or epistemologically inferior to it, reflects an obsolete logical-semantics paradigm which sees language as a code capable of perfectly expressing thought if only it is perfectly used. In a modern, code-plus-inference model of linguistic communication, in contrast, ordinary spontaneous (or even carefully composed) linguistic productions do not perfectly express thoughts and communicative intentions, but merely offer sophisticated evidence for inferring them. In this paradigm, sovereign utterances, as best attempts to represent and communicate 'original' thoughts, enjoy no special status over 'translations', which are best attempts to communicate a thought originating in the act of understanding someone else's utterance; the resulting text or discourse is no less capable of enriching and fertilising a target culture, or stimulating associations in an audience, as the 'original' in its own domain.

Because of their 'necessary evil' status, translating and interpreting attract more pedagogical than scientific interest, since translators must continue to be trained for the foreseeable future. To the extent that interpreting has been studied, the emphasis has also been on the help that cognitive modelling might provide to training. However, quite soon after the emergence of SI, some trainer-researchers who examined recordings and transcripts more closely came to believe that this process of transfer from language to thought to language under special constraints might offer new insights into the relationship between language and mind; and interpreting has also periodically aroused the curiosity of some linguists and psychologists. The question of 'what is so interesting' about SI therefore goes beyond the

elementary questions of how it works or whether it is possible to look at the plausibility of SI as a laboratory for general psycholinguistic research (rather than sociolinguistic, where community interpreting, a less specialised but more intimate and socially-embedded activity, provides a richer seam).

In addition to suffering from being a form of translation, SI is also hard to study, being a live, one-off oral event which can only imperfectly be fixed in audio or written form, and which requires in the researcher near-bilingual language proficiency as well as some grounding in linguistics and/or psychology. It has proved resistant to capture within existing models of language, speech behaviour, attention and so on. But our curiosity should be sparked by the contrast between the scepticism expressed in many quarters at the time of the birth of SI and the satisfactory services it renders daily in dozens of languages and thousands of meetings throughout the world.

At the time of its first spectacular introduction after the second World War, SI was regarded by many as a gimmick and as unworkable for various reasons. The single-channel model of human information processing current at that time ruled out the possibility of attending to two speech streams concurrently. Linguistics was still more concerned with differences between languages than with universals, so it was assumed that lexical and—worse—word-order asymmetries between source and target languages must pose insuperable difficulties: an interpreter working from German or Japanese to English, for example, would always have to wait for the verb. Translation was not normally done in real time and could hardly be accurate at such speeds. Finally, 'lay' interpreters could hardly be expected to understand the technicalities of the meetings and negotiations to be interpreted.

At first, then, like a circus magician's act, SI prompted either scepticism, bafflement or untutored admiration. Scientists, keeping a cool head, offered sober but unfortunately implausible explanations: early modellers assumed, for example, that the interpreter must speak during pauses in the incoming speech, until precise synchronous recordings showed that the two streams were superimposed up to 70-80% of the time. Eventually it was recognised that attention could be shared rather than switched, so that a simultaneous interpreter could be continuously varying her level of attention to various activities—for instance, listening to the incoming speech, making sense of it, searching for expressions and monitoring her own production—depending on the degree of informational density, or conversely, the redundancy or predictability of the material.

The objection based on asymmetrical language structures and the general scepticism about accurate translation at high speeds remained more resistant to counter-argument and demonstration, although in the face of the evidence it is now claimed that conflicting structure does not rule out SI, but must be a source of significant additional cognitive load. This position is indeed more or less entailed in representations of language which stop at syntax and semantics. To resolve the apparent conundrum requires a more 'permissive' model of the flexibility of spoken language, and a revision of an obsolete view of language and meaning that grossly underestimates the role of inference in speech comprehension.

In practice, a trained simultaneous interpreter can produce fluent and reasonably faithful versions of continuous speech with only a short delay because there is enough information from the incoming discourse, when combined with her own relevant knowledge (about the speaker, the event or the subject), either to anticipate outright what the speaker is going to say, or at least to sketch the beginnings of an utterance in such a way as to be able to flesh it out and specify it very soon, as more clues come in, without seeming to stammer or correct herself.

In other words, SI is feasible but only under four conditions: in addition to having the requisite passive and active language proficiency, the interpreter must be *informed* about the

meeting and the affairs of the world, so as to be able to infer enough to fill gaps in the linguistic input; properly *equipped* with access to sound and vision, and thus to the maximum possible range of clues to speaker meaning and audience expectations, with a headset and soundproof booth to isolate the input from her own voice; and *trained* to combine language proficiency, use of knowledge and agile attention allocation on line in these conditions. The steps by which this skill is perfected can be traced in a handful of successful interpreter training schools (Setton, forthcoming), but this is not the topic of the present paper. Here we will highlight two 'interesting' aspects of SI—windows through which it reveals the inferential dimension of language perhaps even more clearly than other instances of language use:

- 'anticipation' as a window on the incremental course of meaning assembly in speech comprehension;

- the use of directive (or 'procedural') devices in discourse by interpreters as a window on how a discourse achieves 'fidelity', or in other words how, beyond mere accuracy, it is made cognitively accessible, to a target audience.

1.   Anticipation: a window on incremental inference

We have already hinted that translation, and perhaps more obviously, simultaneous interpretation, enjoys the dubious privilege shared by many marginal activities of eliciting a strange fascination mixed with distrust. It is here that we shall find the seeds of what makes SI interesting.

The focus of the mixed admiration and disbelief at SI was its speed and simultaneity. A more specific phenomenon which intrigued those who looked more closely at SI transcripts was 'anticipation'. As already mentioned, there were many who doubted the possibility of interpreting simultaneously from a language like German to a language like English because of the final German verb. But lo and behold, here were repeated instances of the interpreter apparently guessing the verb and translating it before it appeared! Here are two examples in SI from German, with a gloss provided in the first, and in the second, a literal translation substituted for the German. In (1), the negated main verb ('needs no introduction') is anticipated (Setton 1999):

(1)   **brauche ich   Herrn O….. - hier –   nicht vorzustellen**
        'need   I   Mr. O….   -   here   -   not   to introduce'
        Int. : ... Mr. O... - he hardly needs any introduction - for as you know he is...

In (2), European railway administrations are discussing the purchase of rolling stock. Here both interpreters anticipate the main verb (and at least one of them also anticipates its object) before either the verb or any of its arguments appear in the original:

(2)   **..if the consortium were to say, let us just suppose they said, we are [in view of this extension**
        Int 1:   ….   and if the consortium said, at
        least let us   suppose that they said
        Int 2:   and if —
        the group -   of manufacturers should then say
        of requests] not prepared [in advance, [i.e.] before the time at which we had
        Int 1:   we do not   agree to - ah

Int 2: the—this is just a hypothesis—that it was not ready   - to <u>provide these</u>
promised to, [the <u>drawings to hand over</u> , then today's discussion is pointless ...
Int 1: <u>give out</u> the drawings earlier   than planned
Int 2: <u>drawings</u> before the time at which they had promised.... (Lederer 1981)

Analysis of such examples seems to lead to the inescapable conclusion—which mainstream utterance-processing theory has now reached by other routes—that online comprehension of unfolding utterances is fine-grained and incremental, and draws on many other sources than simply the linguistic input. First, the formation of representations of meaning cannot hang on one critical constituent, like a main verb, but must be more like the gradual emergence of a photographic image in a bath of developer, or a jigsaw puzzle, in which each piece is not only a structural constituent but also carries a fragment of a picture, evoking other fragments in turn, all of which contribute to building a probable meaning: first a piece evokes a tree, then part of a man's body appears, then a cloud, then a raised foot suggesting that he or somebody in the picture is running. Second, the jigsaws of speech comprehension are formed from both the linguistic items being passed to the jigsaw builder and pieces which he retrieves from other boxes (marked 'other knowledge'), so that a picture may take shape in a variety of routes for different comprehenders on different occasions. The 'cognitive complements' (Lederer 1990) allow provisional representations to be formed and adjusted faster than the linguistic input can spell them out. Trees are inanimate, so 'tree' cannot be the subject of 'run'; clouds often herald rain, from which people may want to shelter, etc. Sense is constructed fast and opportunistically, then adjusted if necessary. In any organism, the use of known patterns to make sense of any input optimises the recognition of threats and opportunities, and in the species which achieved domination by occupying the 'cognitive niche', this reflex continues to drive perception and cognition in the structured dimension of linguistic communication.

Just as our mind fills in the missing parts of an emerging picture from patterns it has seen before, in SI an interpreter completes the sense of an utterance or an idea before the speaker has quite managed to complete the obligatory conventional syntactic articulation, and carried along by her own momentum, expresses meanings as they crystallise within the grammatical constraints and liberties of the target language—which may, for example, mandate early production of a main verb.

Evolved human languages, unlike jigsaw puzzles, are hardly likely to be graded for difficulty. When word order is more flexible, as in German, inflections for case, number, tense etc. often compensate for the delay, telling us in passing the role and circumstances within the picture of the tree, the man, the cloud, etc. This of course does not fully account for the interpreter's often apparently smooth production. Although enough incremental information is being delivered by various clues as the utterance unfolds, the interpreter must produce this information in acceptable sentences (grammatically and communicatively), which is a matter not of language processing theory but of SI techniques, such as formulating in syntactically independent (paratactic) chunks which do not mortgage downstream structure but allow for unexpected changes of course. A reconstruction of the incremental comprehension and production of a German discourse in SI is attempted in Setton (1999: 217 ff.).

Inevitably there will be approximations, later corrected more or less discreetly according to the interpreter's skill. Interpretation differs from translation in the conditions of production perhaps even more starkly than in the conditions of reception of the source text. In addressing their audiences, interpreters can use all the oral resources of the output language such as intonation, rhythm, etc. to compensate for the forced linearity of reception; after all, flexible planning and more or less elegant backtracking and self-correction are common in

everyday conversation. The peculiar, artificially emphatic intonation observed in professional SI, described in Shlesinger (1994), performs several functions. Most importantly, intonation, being suprasegmental, can be used—like parentheticals and 'sentence adverbials' with their loose positional constraints (see next section)—to correct the thrust of a message, even to the point of defeasure, when the interpreter is stuck in an unfolding syntactic construction and contrary evidence comes in, such as (in extreme cases) an unexpected negation. In addition, enhanced voice modulation helps listeners to follow a disembodied voice through a headset; and extra emphasis and explicitation may also help the interpreter herself to identify with what is, after all, someone else's message.

The cognitive processes underlying anticipation are difficult to pin down any more precisely, since words evoke concepts as well as other words (including words in other languages known to an individual), leaving us with a chicken and egg problem until neurologists can reveal that we have either two distinct neural arrays in our brains—concepts and words—or some other arrangement. Lederer (1981) gives examples of prediction based on a simple knowledge of common collocations—for example, adding 'a […] role' when a speaker says 'plays' followed by a long, hesitant series of adjectives. But the distinction she proposes between 'linguistic' and `cognitive' anticipation hardly seems necessary or justified when we do not know whether words are evoking concepts or vice-versa.

The use of external cognitive sources to assemble meaning in SI is attested by the presence in output of elements not semantically encoded in the original, but which are either not noticed, or judged quite relevant and appropriate, by listeners or evaluators. It is misleading to talk of anticipation as a 'strategy'; the process is quite unconscious and concerns all sorts of elements, not just final verbs. In other words, this ostensibly magic phenomenon falls out naturally from, and illustrates, the general principle that inference floods into and fleshes out the piecemeal results of decoding. Sentence processing researchers have recognised that online utterance comprehension cannot possibly wait for the appearance of certain words and syntactic items in the tree, but must be far more fine-grained and incremental, drawing on all sources, including prosodic clues and most crucially, other knowledge or context (Altmann and Steedman 1988).

It will be obvious how anticipation, or prediction, vastly facilitates simultaneous interpreting (and indeed any comprehension process), with no cost to quality provided that a parallel monitoring mechanism is in place to check for unexpected garden paths. The predictability of discourse varies. Chernov (1978 and forthcoming) relates predictability to redundancy, in which he distinguishes 'objective' and 'subjective' components. 'Objective' redundancy is that part which is inherent in language: the fact that words are recognised before they are fully articulated, the unnecessary but mandatory repetition of items like articles and inflections marking case, gender, tense or number, right up to discourse structure in which the tendency of standard communicative discourse is to repeat a former rheme as a theme before introducing a new rheme. All this is further vastly augmented by the 'subjective redundancy' resulting from the interpreter's knowledge about the speaker, situation and previous discourse. This combined redundancy raises an interpreter's (or any hearer's) level of attention to the more global message (or sense) level of the discourse, at which 'probabilistic prediction' enabling anticipation in SI must occur. Chernov showed in experiments how prediction, and therefore viable SI, becomes impossible with very dense or nonsense input— i.e. of very low redundancy—thus ruling out SI of poetry or literature. Subjects were frequently thrown by unexpected continuations, indicating that monitoring of input is switched off, and attentional resources shifted to other operations (self-monitoring or lexical search to enhance style, for example) when a continuation is deemed predictable.

This story of inference, redundancy and triangulation on the message still lacks a principle of selection: how do interpreters, or comprehenders generally for that matter, pick

out precisely the contexts which will maximise the relevance of the utterance? The inputs to the process from an open-ended universe potentially include all knowledge. This 'frame' problem has prompted some eminent cognitive scientists to reject the possibility of modelling higher level cognitive processes altogether, most notoriously in Fodor's 'Principle of Non-existence of Cognitive Science' (Fodor 1983). Relevance Theory (RT) has responded with the Principle of Relevance (Sperber and Wilson 1986/1995). While the wider philosophical and evolutionary underpinnings of this principle are beyond the scope of this paper, a more specific, related RT hypothesis about linguistic communication provides an illuminating backdrop to some other phenomena found in interpretation. The hypothesis is that speakers *guide* hearers to fruitful contexts by means of certain options and devices available in natural language.

2. The pragmatic dimension: connectors and directive devices in SI

The second reason for early disbelief or awe among lay observers of SI was the apparently impossible speed of 'simultaneous translation'. Certainly translating all the words would be impossible. Observing how the whole message can still be conveyed without simplification or significant omission reveals further interesting aspects of SI and of linguistic communication. The elusive items known as 'discourse connectives' in SI transcripts are the visible tip of an iceberg of options and flexibilities of formulation which together make the difference between a communicative—that is to say cognitively accessible—speech and a mere string of encoded propositions. This is the domain of pragmatics, the branch of linguistics which was always loosely concerned with the communicative qualities of language, but has recently been redefined more precisely as the study of inference in communication; in Relevance theory, this now includes the study of those devices which guide inference.

A stubborn implicit assumption of traditional linguistics and some theories of translation, reflected in folk and lay beliefs, was that coding and decoding account exhaustively for what is conveyed from speakers to hearers. This paradigm began to crumble when Austin, Searle and Grice showed how linguistic expressions vastly underdetermine the meanings derived from utterances by hearers, revealing the extent of inference in comprehension. Relevance Theory, updating the Gricean cooperative principle, describes communication as 'ostensive-inferential'. 'Ostension' is the act by which a speaker makes manifest an intention to communicate; in verbal communication this is done by speaking: producing a structured linguistic string which, when decoded, provides a hearer with good *evidence* for what the speaker intends to communicate in a particular situation. Hearers, for their part, process these blueprints in accessible contexts to infer a relevant message.

An interpreter's listeners use inference to derive a message as do those listening directly to the speaker. The constraints on the interpreter's speech to achieve fidelity are therefore not that they should show any particular conventional correspondence to the words or syntax (or phonology or morphology) of the original, but that it should enable TL hearers to derive the same message as the SL hearers get from the original. It is likely that words like 'dehumidifier', 'Tokyo' or '5 million dollars' will be rendered at first mention by fixed conventional equivalents in the TL, but these are limiting cases; as a general rule in translation, the superordinate constraint of fidelity (added in the case of SI to the constraints of incremental online production) results in transformations such that no systematic parallelism can be predicted at any level from morphology through syntax to intonation, and will even be elusive at the level of 'semantics'. Discourse connectives are merely a conveniently visible illustration of the options that can be chosen in TL to achieve this

superordinate constraint of communicative fidelity. Let us look more closely at how this works.

Natural languages are highly sophisticated systems for encoding complex propositions as well as indications of speakers' attitudes, called 'illocutions' in speech-act theory, expressed by means of markers like modal verbs, mood markers, subjunctives, and certain adverbs and particles like *probably, presumably, frankly*. But for most of the specific thoughts and messages which speakers may want to communicate to hearers in specific situations, an additional level of complexity is necessary. Hearers can derive these more complex, situated messages through inference, by combining the output from decoding the linguistic expressions with their existing assumptions and perceptions. But which assumptions (also called contexts) are needed, and how should they be combined with the linguistically explicit meanings? This is the 'ostensive' side of the communication bargain: hearers can only be expected to make useful inferences if speakers help them. To do this, speakers use a variety of linguistic and paralinguistic devices, including intonation, particles such as connectors, and variations on standard word order.

Since hearers have limited attentional resources, communication is most successful when 'contextual effects'[1] are reached by hearers efficiently, i.e. at minimum processing effort for maximum effect.

> The ostensive stimulus must be relevant enough to be worth the addressee's effort to process it; […] at every stage, a hearer should choose the solution involving the least effort. […] An assumption is *relevant* to the extent that its contextual effects in this context are large [and] that the effort required to process it in this context is small; […] information which has the greatest relevance in the initial context [is] that which has the greatest contextual effects and involves the smallest processing effort. (Sperber and Wilson 1986/1995: 127-9; 147; 270-1).

Ostension can therefore be viewed as operating at several levels, from the production of a stimulus to attract attention down to the technical devices which guide hearers as efficiently as possible to effects through the complex structures of human language. The communicativity of a discourse can be evaluated in terms of the efficiency with which it guides hearers to effects. Without language, this is done by expressive gestures and vocalisations; with language, the communicative intention will exploit all the resources of these complex structured vehicles, from prosody through lexical and syntactic choices within the constraints of formal acceptability for a particular language. Since these resources are language-specific, the devices used to modulate speech to optimise communication will vary accordingly. English, a rigid word order language, makes abundant use of intonation; flexible-word-order languages like Czech, Russian and German use variations on word order; some languages, like German, have a rich collection of pragmatic particles for indicating the relative significance of parts of the utterance or signalling that the information contained in it is old, new, secondary, etc. Many languages use cleft word order (*It is this... which... ; lo que no se puede aceptar es que…*) and almost all use parentheticals. Different devices are available in oral speech, where the audio channel is exploited, than in written text; in particular, there are the resources of intonation (pitch, duration, intensity) and rhythm (delivery speed and pauses); and in face-to-face communication, of course, all of which are further complemented with gesture and body language.

Since all linguistic communication uses such signalling devices above and beyond the encoding of the basic information content, and since the range and type of these devices

---

[1] Contextual (or cognitive) effects include any change (addition, modification, substitution, elimination) in the hearer's assumptions, and thus constitute collectively what translation studies conventionally calls 'the message'.

depend on the characteristics of each individual language, we should not be surprised to find that certain words 'appear' or 'disappear' in translation, or that sentences are often restructured in the absence of any apparent strict *grammatical* (well-formedness) constraints.

This can be seen in examples (3)-(5), from the translation into German of an article on interpreting. In (3), the word *auch* (with the rough sense of *even if*) 'appears' in the German translation:

(3) …the interpreter may manage with less knowledge than the participants, since it is sometimes possible to produce language which is transparent to the audience, <u>but not necessarily to the interpreter herself.</u>

…wenn <u>auch</u> nicht unbedingt für den Dolmetscher selbst
... even if not necessarily for the interpreter herself

Conversely, the German particle *ja*, signalling something relevant in the immediate context but also (in contrast to *nämlich*) presumed known to the hearer, may be rendered as *after all* or *as we have seen*, depending on whether the fact is assumed to be in the hearer's long-term memory or was introduced in the previous discourse. Similarly, German *hier* in example (1) above is rendered in the context as *as you know*.

Again, in English (a language evolved among a people fond of irony and understatement), word order and rhythm in examples 4 and 5 may imbue 'as well as' and 'as much as' with an implicit flavour of 'more than', which may have to be rendered more explicitly in another language:

(4) 'speed' of delivery as experienced by a simultaneous interpreter depends on a combination of factors including pausing, intonation, information density and possibly syntactic complexity, <u>as well as</u> speed measured in words or syllables per minute

(5) In dialogue interpreting [in contrast to SI in a booth] the interpreter is directly exposed to the […] emotional overtones of the dialogue, so that the outcome may often depend <u>as much as</u> on management of the interpersonal dynamics of the encounter as on her specialised cognitive skills.

A *communicative* discourse couched in a given language will be shaped, pitched, signposted (or any other analogy one might prefer) in the ways which conventionally draw the attention of native speakers of that language to contexts in which (as RT has it) the decoded semantic product will yield fruitful inferential effects. Such contexts may be present in the hearer's current perceptual field or his short or long-term memory, or indeed, may be about to be introduced by the speaker.

Adapting to its environment of human minds, a viable human language evolves devices which rely on and exploit two kinds of competence in addressees to direct their attention to these sources of contexts:

i. the *cognitive abilities* tacitly assumed in human hearers include awareness of the immediate environment; a short-term 'working memory' in which certain items are in current focus, or remain active long enough to be brought (back) into focus; a long-term memory from which other knowledge can be retrieved; and computational equipment (a deductive device, in RT) which combines whatever assumptions are simultaneously active, delivering modified assumptions in what are known in RT as 'cognitive effects';

ii.   speakers tacitly assume *language competence* in their audience such that particular linguistic devices will bring contexts into focus in which their utterances can be processed to yield cognitive effects.

The linguistic devices which typically direct the hearer's attention to these sources of contexts include, for example, connectors, deictics and definite articles. Cognitive semanticists have described definite and indefinite articles in English as 'pointing to' different areas of memory for entities. Deictics like *soon*, *there*, *this* or *that* 'point to' coordinates of time, place or discourse relative to the speaker's here and now. Sentence-initial subordinating conjunctions like *although*, *if* or *because* signal that the content of the first clause will have to be kept in memory long enough to be processed with the subsequent clause, in addition to signalling aspecific logical relationship between them to the deductive device.

Existing classifications of linguistic items are confusing, falling somewhere between formal and functional criteria. For example, text analysis software sometimes distinguishes between 'content' and 'function' words (in French, *mots signifiants* vs. *mots outils*), assigning adverbs together with nouns, adjectives and verbs to the former category; whereas many 'functional' particles like discourse connectives, phatics and attitudinal and evidential markers are adverbial. Distinguishing 'sentence adverbials' which qualify the whole utterance from those embedded in the verb phrase does not really do the trick either, given the loose constraints on the position of such items in utterances, e.g. *I don't frankly see why she should object*. Ultimately, syntactic or semantic classifications inevitably end up grouping items with widely different roles ('adverbials') but fail to capture functional kinships between items from different formal categories (like '*oh yes'* and ironic intonation, for example).

Relevance theorists have proposed two criteria for a more meaningful classification of linguistic items. The distinction which most concerns us here starts from the intuition that utterances can be expected to encode two types of information, conceptual and procedural, i.e. information about (conceptual) representations, and about how to manipulate them.

> The idea that there are expressions whose function is not so much to encode a concept as to indicate how to 'take' the sentence or phrase in which they occur has played an important role in pragmatics: in particular, in the work of Ducrot and his associates [...]. In speech-act theory, such expressions are treated as illocutionary-force indicators; in the Gricean framework, they are treated as carrying conventional implicatures (Wilson and Sperber 1993: 10).

In fact the roots of this line of inquiry go back at least to the ancient discipline of rhetoric. More recently, theories about the communicative or attention-directing charge carried by certain variations on word order, intonation or rhythm have been developed up to modern times in frameworks like the Prague and Hallidayan schools of functional linguistics and cognitive semantics in the United States.

'How to take the sentence' is clearly too vague and broad a definition of what certain linguistic devices encode. A new analysis has emerged from the study of the elusive category of 'discourse connectors'. Blakemore (1987: 16) noted the difficulty of bringing into consciousness the meaning of items like *well*, *now* or *so* in English, or of learning the meaning of German items like *ja, doch* or *nämlich*, and suggested that these items should be analysed as 'procedural', i.e. of the same nature as grammatical instructions which, unlike words evoking concepts, are processed unconsciously. This led to a principled distinction, spelled out in Wilson and Sperber (1993), between these items and the 'illocutionary' adverbials identified by speech act theory, e.g.

> *Frankly, (confidentially, unfortunately, surprisingly, seriously, sadly)...I can't help you*

Apparently (seemingly, obviously, clearly, possibly, probably)... he couldn't make it.

These items, though traditionally viewed as 'pragmatic', still contribute *explicit* conceptual information ('content'), albeit external to the main proposition, about a speaker's attitude, or the strength of evidence available to him for an assertion. Procedural items, in contrast, like *so, well, after all, anyway, be that as it may*, French *puisque*, Spanish *ya que*, German *ja* and *doch*, certain verbal moods, intonations and word-order variations, and other devices in other languages, are constraints on the *implicatures* to be derived from the utterance. They enhance communicativity by reducing the effort for the hearer, by 'reducing the hypothesis space that has to be searched in arriving at the intended interpretation' (Wilson and Sperber 1993: 21-2) (Articles and personal pronouns are assigned to a different category of procedurals which contribute to the truth conditions of utterances).

Effort, effect and communicativity

Trainers of interpreters are all familiar with a phase, soon after students have mastered a basic set of symbols for consecutive interpreting, when their renditions seem to lack 'cohesive links', and in general a dimension needed to bring the speech to life, above and beyond the simple retailing of propositions, however accurate. In listening critically to tapes of their own renditions, trainees will realise at this point (though not necessarily in these terms) that interpretation is not transcoding, but mind-reading on the basis of linguistic evidence (like ordinary comprehension) and producing evidence from which the audience can derive a fruitful message through inference in various contexts (like ordinary speech). Hearers, interpreters and their audience all *metarepresent* the thoughts of the speakers they follow, with an efficiency which depends largely on the quality of the procedural guidance they are given.

Trainees in translation and interpretation sometimes seem to be recapitulating the process which children go through in learning pragmatic communication, as if progressing through Sperber's (1994) three levels of mind reading sophistication linked to the levels of metarepresentation deployed. In *Naive Optimism*, the hearer assumes the speaker is both communicatively competent and benevolent, so no metarepresentation of his thoughts or communicative intention (as possibly deviating from the decoded surface meaning of his utterance) is necessary. In *Cautious Optimism,* the speaker's competence is not necessarily assumed, so that the hearer may also envisage what the Speaker *might have meant* to convey (for instance, in a slip like *I've been feeding the penguins in Trafalgar Square* (Wilson 2000)). In the third strategy, *Sophisticated Understanding,* the hearer assumes neither the competence nor the benevolence of the speaker (he may be a both a poor speaker, making grammatical mistakes and not finishing sentences, *and* intend to manipulate hearers to his own ends) and may use second-order metarepresentation to infer what the speaker *might have thought the audience would think* was relevant. At first, students of translation and interpretation often seem to approach their texts and speakers at the level of Naive Optimism — expecting them to encode information perfectly and truthfully—whereas *Cautious Optimism* is needed at the very least, and whereas they obviously practice *Sophisticated Understanding* in everyday life.

The interpreter's audience are trying to mind-read as well, so the quality of interpretation will be in part a function of how comfortably, or effortlessly they are helped to do this. Relevance, as a property of utterances, is also a trade-off between the cognitive effects derived and the processing effort required. By reducing effort, 'procedural' devices— discourse connectives, special word order and intonation—as well as lexical choice, play a key role in enhancing communicativity, which is the complement of 'accuracy' (competently

rendering the propositional content, i.e. correctly stating the referents and relations between them encoded in the basic semantic and syntactic structure—most obviously, getting names and numbers right, for instance) in achieving fidelity, the mark of good translation.

Since procedurals must point to contexts as they become necessary to illuminate the unfolding discourse, they must be more intricately intertwined with the syntactic, lexical and prosodic fabric of a language than any other items. Source-language pronunciation and morphology obviously disappear in translation; source-language syntax also to a variable extent, as we have seen. Even conventional lexical-semantic equivalence will be subordinate to the goal of achieving equivalent effect, pointing the hearers to the effects believed to be intended by the speaker. It follows that the 'directive' dimension, since it falls to the most language-specific devices, must be recreated in the target language by an interpreter who has metarepresented the thoughts of the communicator.

This seemingly radical perception, a prerequisite of real communication through interpretation, clashes with many received, learned or 'folk' beliefs about language, and is often only accepted gradually or reluctantly by trainees. A tiny minority of 'naturals' deliver accurate renditions in a convincing and communicative tone almost immediately; another small minority (a problem for teachers) seem to give priority to maintaining personal presence and a persuasive tone at the expense of accuracy while underestimating the difficulty or unfamiliarity of the subject matter. The majority of trainees follow a typical learning curve in which for the first few months of the course, they fall below their normal standards of coherence and communicativity in speech production while grappling with translation-specific problems and unfamiliar subject matter, only recovering their communicative abilities after various techniques and strategies have been mastered. This development can be explained in terms of cognitive resource allocation: while acquiring new task-specific procedures, like note-taking for consecutive or the peculiar pattern of online attention management required for SI, they neglect the self-monitoring which usually accompanies normal speech; once the new procedures are stabilised and internalised, the interpreter again becomes aware of her own discourse and can put herself in the place of' her audience.

Interdisciplinarity and the potential of T & I research

In the short history of research on interpreting, interest in the activity as a window on language and thought has been a by-product, at best; the main effort has been geared to the needs of pedagogy and the related issues of quality and certification. An early radical-pragmatic approach emphasising communicativity and favouring corpus studies (the 'Paris school'), dominant until the late 1980s, has now largely been displaced by an experimental psycholinguistics paradigm focussing on the cognitive modelling of human information processing. Information transfer is deemed more manageable than the more elusive determinants of communication, which are bracketed off for 'qualitative' research as aspects of 'presentation' or 'style'. Although the proponents of the experimental paradigm have been careful not to raise expectations, its harvest of findings or insights has been disappointing—predictably so, since it is no less challenged than its predecessors and competitors (introspective or discourse-analytic methods, for example) by the multiplicity of variables and the elusiveness of a phenomenon as complex as situated cross-linguistic verbal interaction.

We may indeed baulk at trying to model a process as complex as SI, when others are patiently and modestly modelling the comprehension of simple *sentences*. However, it is now accepted in science that woods may also teach us something about trees, that certain patterns may emerge only beyond a certain threshold of complexity, and that we may learn as much

from observing complex systems in operation *in situ* and *in toto* as by dissecting them into supposedly more elementary and tractable components.

What methods might help us consolidate our theories about communicativity? After all, students are also often careless about including ordinary illocutionary markers of attitude, evidence or discourse orientation like *I think, personally, strictly speaking*, *in conclusion*, *furthermore* or *to my mind,* and all interpreters are notoriously vulnerable to error on the most 'core semantic' items like numbers and proper names. Is there a criterion and an experimental design that could help isolate items which contribute to communicativity in the precise *procedural* sense we have described ?

Pragmaticists and translators share a pronounced awareness of the all-pervasive role of context in linguistic communication. Until recently, however, translators - perhaps due to too close an involvement in the reality of the interpretation and production of communicative discourse – have not been able to make a clear and perspicuous contribution to theory. The cognitive (as opposed from sociological) side of contemporary translation studies relies a good deal on introspection, albeit of a sophisticated kind using think-aloud protocols, etc.). Introspection nevertheless remains a 'soft' methodology, and is hardly workable for SI. In the heat of the action, professional interpreters and translators may not even notice or experience any difficulty in rendering utterances containing procedural or directive devices. One possibility might be to explore the hypothesis that procedurals correlate with the least readily 'translatable' items in language (the other pole being represented by technical and single-referent terms like 'dehumidifier' or 'NATO'). Professionals working into a range of target languages to grade a list of selected items in English, for example, on a scale of 'context-independent' translatability, or the confidence with which they could propose a reliable all-context equivalent.

A variety of indirect methods are conceivable: one might use judges to evaluate the effect—helpful, clarifying, distorting or negligible—of the spontaneous addition or 'omission' of certain items in translation, and look for a correlation with items classified in relevance-theoretic work as procedural or conceptual; compare interpretation performances, or the quality of listener reception (on the basis of synopses which listeners would be asked to write of the speech) from discourses with and without 'lubricant' procedural devices.

Corpus-based research

One possible avenue for the future study of fidelity, for purposes pedagogical or sociological, might lie in corpus-based research, in which interest is now being revived with the recent access to much larger corpora (from the European Parliament, for example) and the help of text analysis software. Analysis of such natural data must allow for the loss of most of the original context, even with live recordings and supporting documentation, and will of course always be dependent for its framework on basic theory in psychology, linguistics and pragmatics. For quantitative research, large representative corpora seem essential to credibility, but automatic measurements will only produce nonsense without some basic understanding of the ecological conditions in which interpreting is set. In approaching a corpus we can make three kinds of legitimate assumption about the interpreter:

1. Motivational: she is doing her best to experience and express the speaker's meaning;
2. Cognitive: that her formulation at time t can only use what assumptions she has available about the speaker's communicative intention at time t —whether from speech input, memory or current perception, and consisting to a large extent of inferences (some of which the speaker may go on to express explicitly, hence, 'anticipation').

3. Linguistic constraints: she must produce conventionally acceptable speech, and since languages differ and speakers are less constrained than interpreters in this regard, she usually has to build her own syntax as she goes along, favouring 'chunked', paratactic constructions to allow for possible changes of course.

Together these assumptions give us a basis for observing the interpreter's memory, her self-imposed norms, and so on. There are broadly two ways of analysing a corpus. One method, used to highlight the phenomena described in this paper (Setton 1999), is to produce a synchronised record in which to compare input and output locally and thus study memory, lag, online SI strategies and so on depending on the detail provided. Microanalysis of this kind can focus on a class of items (verb-final sentences, for example, or parentheticals) or aim to reconstruct the process as a kind of time-and-motion study, perhaps attempting to infer cognitive processes. One can compare what is expressed in production at time t to the baseline semantic content of the previous incoming utterance(s), extending back, say, 12 seconds (the maximum lag reported in the literature[2]) and determining the source of what is left. When we do this, we find plenty of material sourced from long ago in the discourse or externally to the discourse; so much so, in fact, as to suggest that ear-voice span (EVS), defined as the 'time from an item in SL to production of its equivalent in TL' can probably only be reliably measured for certain items with stable equivalents (like proper names or numbers) and therefore has little value except as an indication of an interpreter's general strategic lag.

In looking for overall quantifiable patterns in large, representative corpora, the same modesty and controlled expectations are called for as in the experimental paradigm. However plausible modern pragmatic theories might be, to define and unravel clear quantifiable entities remains difficult and delicate work. This is illustrated in a recent and ongoing project to test the hypothesis that interpreting quality as judged by frequent users of the service correlates with restructuring and free supplying of connectives by the interpreter (Setton and Motta, in preparation). These (independent) variables are themselves difficult to define; and the picture is complicated further by subjects who, like the second category of students mentioned earlier, may use connectors and similar items abundantly as padding or to cover up for inadequate comprehension.

Conclusion

What contribution can the study of T & I make, and in what sense must it be interdisciplinary? Given the important role interpreting and translation play in cross-cultural communication and international life, it is not unreasonable to hope that research into these activities might help us better to understand the limits on communication across languages and cultures, making a contribution to ethnology, sociology or anthropology. Numerous research projects have explored distortion, dilution, inaccuracy and even ideological bias in translation and interpretation. On another level, studying interpreting or translation as operations of individual minds, albeit of a peculiar bilingual sort, may offer insights into some limits and unexpected potentialities of human cognition, specifically into the relationships of (in)dependence between thought and language(s); lastly, targeted empirical studies may also simply inform the translator training effort. These are potential dividends for psychology, cognitive science and education.

How realistic is the prospect of extracting real knowledge from the study of translation and interpretation? Whether the emphasis is sociological or cognitive, it will be clear by now that the object of study is one of bewildering complexity, and as in the study of any human

---

[2] Oléron & Nanpon 1965; Lamberger Felber 2001.

82

behaviour, is further obscured by reflexive interference (we are studying ourselves), so much so that in many cases informed intuition still seems more reliable than sophisticated scientific methodology in picking out significant patterns. No-one, for example, contemplates automating any part of the role of juries in translators' or interpreters' qualifying examinations, or even deferring to a model or theory.

This challenge of complexity is illustrated in the two SI phenomena which we have analysed in this paper in terms of an inferential model of communication. We see inference as the key to the very possibility of SI: interpreters 'leverage' the inferential dimension to circumvent and offset both the formal differences between the source and target languages and the inevitable gap in knowledge between themselves and participants at specialised meetings. But the methods used soon come up against limits: we can only say that anticipation strongly indicates a certain pattern of utterance processing; and even with the help of recent work in pragmatics, we can only begin to circumscribe the set of features which makes discourse communicative. In addition to tangible items like 'discourse particles', that set probably includes a wide range of features achieving similar effects, like special word orders, connoted lexical choices and intonation contours, possibly shading off into still more elusive combinations of rhythm, tone and so on which are increasingly difficult to fix, record or describe. If this project seems elusive and daunting today, we should remember that most of scientific progress in the last two hundred years has involved devising ways of indirectly observing and measuring phenomena which could not be observed directly—elementary particles are a prime example. Let us hope that the linguistic correlates of communicativity will be less elusive.

Robin.Setton@eti.unige.ch

References

Altmann, Gerry and Steedman, Mark. 1988. "Interaction with context in human sentence processing." *Cognition* 30, 191-238.

Blakemore, Diane. 1987. *Semantic Constraints on Relevance*. Oxford: Blackwell.

Chernov, Ghelly V. 1978. *Theory and Practice of Simultaneous Interpretation.* Moscow: Mezhdunarodnyie otnosheniya. (In Russian).

Chernov, Ghelly V. (2004) *Inference and anticipation in simultaneous interpreting.* Critical edition with foreword and notes by Robin Setton and Adelina Hild. Amsterdam/Philadelphia: John Benjamins.

Fodor, Jerry. 1983. *The Modularity of Mind*. Cambridge, Mass.: MIT Press.

Keenan, Edward. 1978. "Some logical problems in translation." In F. Guenthner and M. Guenthner-Reutter (eds.), *Meaning and Translation: Philosophical and Linguistic Approaches,* London: Duckworth, 157-189.

Lamberger-Felber, Heike. 2001.Text-oriented research into interpreting: examples from a case-study. *Hermes* 26, 39–63.

Lederer, Marianne. 1981. *La traduction simultanée*. Paris: Minard Lettres Modernes.

Lederer, Marianne. 1990. "Cognitive complements in interpreting." In D. Bowen and M. Bowen (eds.), *Interpreting, Yesterday, Today and Tomorrow*, American Translators Association Scholarly Monograph Series, Vol. IV. New York: SUNY, Binghampton.

Oléron, Pierre and Nanpon, Hubert. 1965. "Recherches sur la traduction simultanée." *Journal de psychologie normale et pathologique* 62, 73-94.

Quine, Willard V. 1960. *Word and Object*. Cambridge, Mass: MIT Press.

Setton, Robin. 1999. *Simultaneous interpretation: a cognitive-pragmatic analysis.* Amsterdam/ Philadelphia: Benjamins.

Setton, Robin. (forthcoming). "New demands on interpreting and the learning curve in interpreter training." Paper delivered at the Fifth National Conference on Interpreting Practice, Pedagogy, and Research, Shanghai, November 6-7, 2004.

Setton, Robin and Motta, Manuela (in preparation). "Syntacrobatics: restructuring and formal autonomy in simultaneous interpretation with text".

Shlesinger, Miriam. 1994. "Intonation In the Production and Perception of Simultaneous Interpretation." In Lambert, Sylvie. and Moser-Mercer, Barbara (eds), *Bridging the Gap: Empirical Research in Simultaneous Interpretation*, Amsterdam/Philadelphia: John Benjamins, 225-236.

Sperber, Dan. 1994. "Understanding verbal understanding." In Jean Khalfa (ed.) *What is Intelligence?* Cambridge: Cambridge University Press, 179-198.

Sperber, Dan and Wilson, Deirdre. 1986/1995. *Relevance: Communication and Cognition*. Oxford: Blackwell.

Wilson, Deirdre. 2000. *Metarepresentation in linguistic communication*. In Dan Sperber (ed.), *Metarepresentations: A Multidisciplinary Perspective*. Oxford: Oxford University Press. 411-448.

Wilson, Deirdre and Sperber, Dan. 1993. "Linguistic form and relevance." *Lingua* 90 [Special Issue on Relevance Theory, Vol. 2, D. Wilson and N. Smith (eds.)], 1-24.