Grammaticalization as the Fossilization of Constraints on Interpretation: Towards a Single Theory of Communication, Cognition, and the Development of Language*

Randy J. LaPolla
City University of Hong Kong;
Institute of Linguistics, Academia Sinica
ctrandy@cityu.edu.hk

ABSTRACT

Combining insights from work on Relevance Theory, contextualizations cues, 'emergent grammar', and grammaticalization, this paper argues that inference is the basis of all communication, and of all aspects of linguistic communication, and that grammar is a set of evolved social/cultural conventions that aid the inferential process by constraining that process. Morphological forms develop as speakers repeatedly use lexical items with conceptual information in contexts where the conceptual information is not necessary to the main information of the utterance, and the form is used simply to constrain the hearer's search for the relevance of the utterance, such as making the relationships between various parts of the total message more explicit or situating the message temporally. A fully grammaticalized form is said to contain only procedural information, but all grammaticalized forms develop from full lexical forms used for marking procedural information, and this is what allows them to become grammaticalized. Syntactic patterns become grammaticalized as implicatures and associated aspects of particular repeated patterns of use (such as same subject patterns in coordination) become fixed. These then can aid in referent identification and other aspects of interpretation. The development is not teleological, but evolutionary. Languages then come to differ in terms of what aspects of the interpretation are obligatorily constrained and what are not, and in terms of the degree to which interpretation of a particular functional domain is constrained. Languages also differ in the way they constrain interpretation of a particular functional domain.

1. Introduction. This paper is about communication and how it gives rise to, and influences, the development of language structure. The discussion will be mainly from the point of view of Relevance Theory, but we will see that several independent lines of inquiry in linguistics have involved concepts parallel to those in Relevance Theory, and these concepts, as well as certain others I will mention, can be combined with those of Relevance Theory to form a more comprehensive theory of communication, cognition and the development of language structure.

2. Communication
2.1. Relevance Theory. Relevance Theory (Sperber & Wilson 1986/1996) is a theory of communication and cognition. From the point of view of Relevance Theory, communication involves two information processing devices (e.g. two people). One does something (an ostensive act) to cause the other to come to share some information. Interpretation of

---

linguistic communication is said in Relevance Theory to involve two types of process: the decoding of a linguistic expression, which produces a LOGICAL FORM (a blueprint of meaning), followed by an inferential process which first completes the logical form, transforming it into a full proposition expressing the explicit content of the utterance (what is called the EXPLICATURE, a concept similar to what was called the 'sentence meaning' by Grice), and then derives any implicatures that must be created in the processing of the explicature for the use of the explicature in that context to 'make sense' (what Grice called the 'speaker meaning'). Interpretation then, starting with the filling out of the explicature, involves more than simply decoding a signal. The pragmatic aspects of meaning, such as resolving ambiguities, identifying referents, identifying illocutionary force, recognizing irony and humor, and completing incomplete utterances, all must be interpreted by inference. The inference involved in communication is essentially guesses at what the communicator's intended message might be. These guesses are possible because of the unconscious assumption of the principle of relevance, given in (1).

(1) The principle of relevance (Sperber & Wilson 1996:260/270)
      (1) Human cognition tends to be geared to the maximisation of relevance.
      (2) Every act of ostensive communication communicates a presumption of its own optimal relevance, such that
          (a) The ostensive stimulus is relevant enough for it to be worth the addressee's effort to process it.
          (b) The ostensive stimulus is the most relevant one compatible with the communicator's abilities and preferences.

The speaker then will tailor the utterance, in the case of linguistic communication, in such a way that the hearer will not have to expend unnecessary effort to create a context that will allow him/her to arrive at the intended interpretation. In doing this, the speaker takes into consideration guesses as to what information is available to the hearer at the time of utterance for use in interpreting the utterance. A speaker must decide what to make explicit and what to make implicit (and also, among implicatures, what to make stronger or weaker implicatures), and this is done on the basis of the speaker's estimation of the hearer's processing abilities and contextual resources, but also partly on politeness considerations and what we think of as 'style'.[1] The more information that the speaker assumes the hearer is able to access in the processing of an utterance, the less explicit the utterance can be. For

---

[1]An example of the use of a particular utterance form in order to convey weak implicatures for the sake of politeness is the following:

      A: Would you like to go see a movie tonight?
      B: Thanks, but I have an important test tomorrow morning.

The implicated conclusion, a strong implicature, is that B cannot go to the movies that night, but there is a weak implicature that were it not for having that test the next day B WOULD go with A to the movies, and it is in order to convey this weak implicature that this particular form of utterance is chosen. (The proposition conveyed by the weak implicature need not be true; it may be that the speaker is just trying to be polite (save A's 'face')—the strength of an implicature is directly proportional to the degree to which the speaker takes responsibility for the hearer making that particular interpretation. In this case, if B wanted to be sure A made that interpretation, B could add *How about next week*, or some such expression.)

example, in the Philippines I once saw a sign advertising 'Owner for sale'. The person who made this sign was assuming that the intended addressees, in constructing the set of assumptions for processing this expression, will have access to the assumption that there are two types of jeepnies (a modified jeep), one called 'owner-jeepny' (owner-operated, for personal use) or simply 'owner', and one called 'passenger-jeepny' (used commercially for carrying passengers), and that since it doesn't make sense to say 'owner for sale' unless 'owner' has some meaning in this context other than 'one who owns', the addressees will use that assumption in processing the expression, and arrive at the intended interpretation, that an owner-operated jeepny is for sale. If instead the target audience was to include people not familiar with this culture, the sign-writer would need to be much more explicit, adding much more lexical information to constrain the interpretation, for example writing 'Owner-operated (non-commercial) jeepny for sale'. When necessary because of the speaker's estimation of the hearer's processing abilities and contextual resources, parts of an utterance may be produced solely to assist the hearer in interpreting the main message of the utterance. There are often a great number of degrees of explicitness possible, depending on the speaker's estimation of the hearer's inferential abilities and current knowledge state; the more explicit the utterance, the more constrained the interpretation, as in the six different possible answers to the question given in (2) (all of which have the same 'meaning'; of these, the first is attested).

(2)   Q: Do you want something to drink?
       A1:   (points to soup bowl)
       A2:   I have soup.
       A3:   No. I have soup.
       A4:   No, because I have soup.
       A5:   No, since I have soup, I don't need anything to drink.
       A6:   No, I don't want anything to drink. Since I have soup, I don't need anything else to drink right now.

The degree to which the hearer is forced to deduce a particular implicature depends on the degree to which the form of the utterance constrains the hearer in choosing the contextual assumptions necessary to achieve relevance in interpreting the utterance. Aside from adding more conceptual information, that is, more lexical items, as in the jeepny example, it has been argued in Relevance Theory (Blakemore 1987, 1988a,b, 1990; Wilson & Sperber 1993; Nicolle 1997) that one way the speaker can constrain the interpretation of implicature is to use discourse connectives such as *so* and *after all*, which are said to contain procedural information (procedures for manipulating conceptual representations), that is, information on how to interpret the proposition, to alert the hearer to the fact that one part of the utterance has a particular relationship to another part of the utterance, such as providing additional evidence or an explanation. In (3a-b) is an example with two possible interpretations (from Wilson & Sperber 1993:11). In one interpretation the statement in (3a) provides evidence for the conclusion in (3b); in the other the conclusion in (3a) is confirmed by the evidence presented in the statement in (3b). In this case it would be possible for the speaker to constrain the hearer's choice of one or the other of these two interpretations by adding either *so* or *after all* to the beginning of the second clause, as in (4a) and (4b) respectively.

(3)   (a) Peter's not stupid. (b) He can find his own way home.
(4)   a.   Peter's not stupid; *so* he can find his own way home.
   b.   Peter's not stupid; *after all*, he can find his own way home.

The connectives *because* in (2A4) and *since* in (2A5) and (2A6) above perform a similar function in making the relationship between the two propositions explicit, thereby constraining the interpretation. Discourse connectives such as these are said to not encode concepts (that is, they do not contribute to truth conditions); they just constrain the inferential phase of the comprehension, narrowing down the search for relevance, and thereby make the search easier, and make the interpretation selected more determinate.

  2.2. Contextualization cues. There is a parallel line of research within conversation analysis that has also discussed constraints on the process of interpreting communicative acts. This is the work done by John J. Gumperz on contextualization cues (e.g. Gumperz 1977, 1982, 1989, 1992a, 1992b).   For Gumperz, interpretation is based on what he calls 'conversational inference', which is 'the situated or context-bound process of interpretation, by means of which participants in an exchange assess other's intentions, and on which they base their responses' (Gumperz 1982:153). Here too relevance is seen as the key factor, as interpreting an utterance is said to involve 'a search for an interpretation that makes sense in terms of what we know and what we have perceived.' (Gumperz 1977:204).   Hearers interpret the meaning of an utterance based on inferences about the speaker's underlying strategies and intentions, and these inferences are drawn on the basis of interpretive frames (contexts) evoked by certain linguistic or non-linguistic contextualization cues produced by the speaker. An utterance might have many possible interpretations, and the hearer decides how to interpret the utterance based on his or her understanding of what is going on in the interaction (what type of activity the interaction is), and this understanding itself is based on recognition of identifiable or familiar patterns (frames) (Gumperz 1982:130). The type of frame 'does not determine meaning but simply constrains interpretations by channeling inferences so as to foreground or make relevant certain aspects of background knowledge and to underplay others' (Gumperz 1982:131). The speaker signals the type of activity, and alerts the hearer to how the semantic content is supposed to be interpreted and how the particular utterance relates to those that precede or follow. This is done using contextualization cues, which are elements of the utterance that help to signal contextual presuppositions (Gumperz 1982:130), essentially what we talked about in the discussion of Relevance Theory above as 'procedural information'. Gumperz argues that while these contextualization cues convey information, the information is not like that of lexical items, as it is independent of the propositional meaning, and conveyed only in the interactive communicative process, and so cannot be discussed outside of that process.

  An example of how prosody, one type of contextualization cue, can influence how a hearer processes an utterance is the use of a rising tone to signal the relationship between two clauses. In (5) (from Gumperz 1977:201), if the first clause ends on a rising tone, then it will be interpreted as subordinate to the second clause, and therefore the relation between the two clauses will be understood as cause and effect. But if instead the first clause ends with a falling tone, then the link between the two clauses would be difficult to make, and so the interpretation might not be of a cause-effect relationship between these two clauses,

especially if there were another clause preceding the first clause with which it might be linked instead, such as *I can't talk to him now*.

    (5)   Because I'm busy / I don't want to be interrupted //

Another example is the difference in interpretation between (6a) and (6b) (from Gumperz 1982:110):

    (6)   a. My sister who lives in New York / is very nice //
            b. My sister / who lives in New York / is very nice //

Here the difference in tone grouping is one of the main cues used to distinguish whether the phrase *who lives in New York* is a restrictive or a non-restrictive relative clause. (See also Halliday 1994:295-97 for similar examples.)

    Intonation then can have a similar function to the discourse connectives discussed above in constraining the search for the most relevant interpretation of the relationship between two clauses. Gumperz also talks about how prosodic factors are used to determine higher level implicatures, such as whether to interpret an utterance as a joke or as sarcasm, etc (see also Selting 1992). In (7a-b) (from Gumperz 1982:110), the difference in tone grouping leads us to interpret (7a) as a simple clarification, but (7b) as conveying an attitude of annoyance or impatience.

    (7)   a. I said sit down //
            b. I said / sit / down //

    Accent placement is also a key cue in interpretation. It marks new from old information (i.e. topic from focus), and links different parts of the text together, allowing us to trace the thematic development of an exposition (Gumperz 1982:114; see also Gumperz, Aulakh, and Kaltman 1982).

    Gumperz has done quite a bit of work looking at how these cues work cross-linguistically, and has shown how not using the correct cues for a particular culture often leads to misunderstanding. Use of one's native cues when speaking a second language occurs because these patterns of language use are so strong that they carry over to the second language, particularly if the speaker is unaware of the possible differences of interpretation of the cues in the other culture. One attested example Gumperz gives (1982:126; citing Mishra 1980) is when an Indian English speaker talking to a Western English speaker used extra emphatic stress in saying (8) and (9):

    (8)   In the third school / in which I "had been transferred.
    (9)   So I went to see the / . . . an"other union person.

A Western English speaker is likely to arrive at the interpretations given in (10) and (11) respectively:

    (10) In the third school to which, in spite of what you might think, I had really been transferred.

(11) So I went to see another person from the union.

Yet the interpretations intended by the speaker, and also given by other Indian speakers of English interviewed by Gumperz, were those in (12) and (13):

(12)   In the third school, to which I was transferred against my will.
(13)   So I went to see a person from a different union.

The difference in interpretation is due to the difference in the culturally determined understanding of the use of extra-emphatic stress.

For Gumperz, contextualization cues are involved in all aspects of interpretation:

> To the extent that our knowledge and use of contextualization cues is a function of shared interactive history and rests on socially based presuppositions, we can say that social knowledge is part of the input that determines what we perceive as linguistic reality. In other words, all perception and interpretation rests on selective processing of perceptually available cues. Contextualization conventions help to determine which of such cues are seen as information bearing in the first place. (1992b:50)

More specifically, contextualization cues can assist in three distinct levels of interpretation involving contextualization-based inferential processes. At the most basic level the perception of auditory and visual communicative signals and their analysis into information units involves inference guided by contextualization cues, and is not a simple decoding as assumed in Relevance Theory (Gumperz 1992:232). At the second level, the level of speech act implicatures (or sequencing), contextualization cues assist in the determination of what Gumperz calls the 'communicative intent'. At this level are included inferences not explicitly coded in the lexical content (Gumperz 1992:232-233; cf. Wilson & Sperber's (1993:5) 'higher-level explicatures'). At the highest, most global level of framing, contextualization cues can give rise to expectations about what is to happen in the communicative activity and resolve possible ambiguities at the perceptual or speech act implicature levels (Gumperz 1992:233). Gumperz argues that '[c]ontextualization cues on this view can simultaneously act as local level signals and serve as inputs to higher level processes of conversational inference which retrieve the cultural presuppositions about the activity with reference to which interpretations are made and validated' (1989:78-9). The levels discussed by Gumperz as the local vs. higher levels are essentially what in Relevance Theory is called explicature (and higher-level explicature) vs. implicature.

Gumperz has focused his discussions on the use of prosody; deixis and anaphora; lexical, code and style choice; use of certain speech-act verbs and formulaic expressions; and what Gumperz calls 'paralinguistic signs of tempo, pausing and hesitation, conversational synchrony ... and other "tone of voice" expressive cues'. But the term 'contextualization cue', according to Gumperz (1977:199), 'refers to any aspect of the surface form of utterances which, when mapped onto message content, can be shown to be functional in the signaling of interpretive frames'. In Gumperz, Aulakh, & Kaltman 1982, examples are given of how particular uses of word order, particular inflections of the verb, emphatic particles, focus marking particles, conjunctions, deictic marking, and repetition can all be involved in

the signaling of interpretive frames (see also Uhmann 1992 on the contextualizing function of speech rate changes).

2.3. Interim summary.   We have seen in the sections just presented that two completely independent lines of research, one based in philosophy, one in anthropologically oriented conversation analysis, have arrived at the idea that communication is based on relevance-driven inference, and that this inference is constrained by aspects of the linguistic or even non-linguistic form of the communicative act.   Relevance Theory theorists have talked mainly about certain clause-linking particles as constraining interpretation, as having procedural rather than conceptual content, while Gumperz has mainly given examples of prosodic factors used as cues in interpretation.   Relevance Theory accepts the Chomskyan view of autonomous syntax, and so inference begins after the recovery of the logical form, while for Gumperz all aspects of interpretation, including the recovery of the form of the communicative act, involve inference.

We will come back to these issues, but first we will look at some of the differences in the ways languages can constrain interpretation.

3. Grammar as constraints on the search for relevance.   For a number of years I have been arguing that Chinese and most other Sino-Tibetan languages do not work the same way, in terms of pivots and grammatical relations, as either languages with largely nominative-accusative structure, such as English, or those that have largely ergative structure, such as Dyirbal (LaPolla 1988, 1990, 1993, 1995a, 1996). For example, in a language with an [S,A] pivot for coordination (the accusative pattern), such as English, an argument shared by two conjoined unmarked transitive clauses can be represented by a zero pronoun in the second clause only if it is the actor (A role argument) in both clauses, as in (14a).

(14) a.   The man went downhill and Ø saw the dog.
      b.   *The dog went downhill and the man saw Ø .
      c.   The dog went downhill and Ø was seen by the man.

It is not possible to have the representation of the actor of the first clause coreferring with a zero pronoun representing the undergoer (P role argument) of the second clause without using a passive construction, as shown in (14b). It is not possible to say *The dog went downhill and the man saw*.   If the argument the two clauses have in common is the undergoer of the second clause, in order for the two clauses to be conjoined, the representation of the argument (here the zero pronoun) must appear as the single direct argument of a passive construction, as in (14c).

In a language with an [S,P] pivot for coordination (the ergative pattern), such as Dyirbal (Dixon 1980:461ff), a shared argument which appears as a zero pronoun in the second of two conjoined clauses must be in the S or P role in each clause, as in (15a). If the argument in the second clause is instead in the A role, in order for the two clauses to be conjoined and for the argument to be represented by a zero pronoun in the second clause, the shared argument must appear as the derived S of an antipassive construction, as in (15b). It is not possible to say

the equivalent of *The man went downhill and saw the dog* with a transitive second verb and a zero anaphor referring to an A argument, as in (15c) (from Dixon 1980:461-2.).[2]

(15) a. balan   guda   buŋa-n   baŋgul yara-ŋgu bura-n
        she+ABS dog+ABS descend-PAST he+ERG man-ERG see-PAST
        The dog went downhill and was seen by the man (Lit.: The dog went downhill and the man saw Ø.)

    b. bayi   yara   buŋa-n   bulralŋanyu   bagun gudagu
        he+ABS man+ABS descend-PAST see+PAST+ANTI he+ABS dog+DAT
        The man went downhill and saw the dog (with antipassive indicator ŋa-y on the second verb).

    c. *bayi   yara   buŋa-n   bura-n   baŋgul guda
         he+ABS man+ABS descend-PAST see-PAST he+ERG dog+ABS
        The man went downhill and saw the dog (with transitive verb and A argument (yaraŋgu) unexpressed).

In Chinese we don't find either the English or the Dyirbal type of restriction on cross-clause coreference. In Chinese it is possible for the shared argument of a conjoined structure to be deleted regardless of whether it is in the A or P role, as we can see from the examples in (16):

(16) a. Xiǎo gǒu$_i$ zǒu dào shān   dìxià, nèi ge rén   jiù kànjiàn le Ø$_i$.
        little dog walk to   mountain bottom that CL person then see   CSM
        'The little dog went downhill and was seen by the man.'
        (Lit.: 'The little dog went downhill and the man saw Ø.')

    b. Nèi ge rén$_i$ zǒu dào shān   dìxià, jiù Ø$_i$ kànjiàn le   xiǎo gǒu.
        that CL person walk to   mountain bottom then   see   CSM little dog
        'The man went downhill and saw the little dog.'

The result of this situation is that in languages with grammatical constraints on the control of anaphor like those we've just looked at, those constraints force a particular interpretation of a sentence, even if the result is nonsensical, as in (17), from Comrie (1988:191):

(17) The man dropped the melon and burst.

Because of the grammatical constraint on conjunction reduction in English, this sentence has to be interpreted as saying that the man burst after dropping the melon. That is, when there is a coordinate structure such as this, the rules of English syntax force the interpretation that the zero pronoun is coreferential with the S or A role argument of the first clause, and block the

---

[2]Abbreviations used in the examples: ABS absolutive, AGT agentive; ANTI antipassive, CL classifier, CSM change of state, DAT dative, DIR directional; ERG ergative, LOC locative; INF inferentially derived conclusion; IP third person intransitive past; NP non-past declarative, PF perfective; PROG progressive, PS predicate sequence; R/M reflexive/middle, TMdys past tense, 1 day-1 year ago; TMhrs past tense, within today; TMyrs past tense, years ago; TP 3rd pers. transitive past.

semantically more likely interpretation that it is the O role argument that controls the zero anaphor, that it is the melon that burst. In a language such as Chinese, though, where there is no such grammatical constraint on interpretation, the equivalent sentence would not force such an interpretation, even with the man being the topic of the utterance, as real world semantics would influence the interpretation more than the structure. Over the years I have asked well over a hundred native speakers of Chinese to translate this sentence into Chinese and then tell me who or what burst. The answer is invariably 'Of course the melon burst.' They are generally quite surprised when I tell them that the English sentence MUST mean that the man burst.

In Rvwang, a Tibeto-Burman language spoken in Northern Burma, we have the same lack of constraints on the interpretation of clause coordination, as evidenced by the pair of sentences in (18).

(18) a. Vpūngí Vdǿsv̀ng vdip bǿà nǿ ngǿa:pmì
          Vpūng-í   Vdǿ-sv̀ng   vdip bǿ-à   nǿ ngǿ-ap-ì
          Apung-AGT Adeu-LOC hit   PF-TP PS cry-TMdys-IP
          'Apung hit Adeu and cried.' (Adeu cried)

      b. Vpūngí Vdǿsv̀ng vdip bǿà nǿ vhø̄shì a:pmì
          Vpūng-í   Vdǿ-sv̀ng   vdip bǿ-à   nǿ vhø̄-shì   ap-ì
          Apung-AGT Adeu-LOC hit   PF-TP PS laugh-R/M TMdys-IP
          'Apung hit Adeu and laughed.' (Apung laughed).

Here the structures are exactly the same, though the actor of the second clause is interpreted differently due to real world expectations of who would be more likely to cry or laugh after an act of hitting. In fact the interpretation is quite unrestrained; although I've written 'Adeu cried' and 'Apung laughed' after the free translations, actually the interpretation could be that the one who cried or laughed was either one of these two people, or even a third person, such as someone standing nearby watching what was happening between Adeu and Apung. Most Sino-Tibetan languages are similar to Chinese and Rvwang in not having syntactic constraints that force particular interpretations of cross-clause coreference.

Let's look at some other ways that the grammar of English constrains interpretation. One way is with verb agreement. Aside from the obvious effect that verb agreement has on the identification of particular arguments, it can also constrain the interpretation of the syntactic structure. To borrow an example from Georgia Green (1996:144), the use of singular vs. plural agreement in (19a) and (19b) forces two different analyses of the structures. In (19a) pickles and ice-cream must be interpreted as two different items about which the same predication is made, while in (19b) they must be interpreted as one item (a dish with two things combined) about which a predication is made.

(19) a. Pickles and ice cream are really great.
      b. Pickles and ice cream is really great.

In Chinese it is not possible to constrain the interpretation in this way, as there is no agreement marking, so there would be only one form for both these meanings in Chinese; the

inferential process involved in deciding on the proper structure (and therefore the proper interpretation) would not be constrained by the linguistic form in the way that it is in English.

Tense marking also restricts the search for the relevant interpretation. For example, to interpret the proper time frame for the situation expressed by the Chinese sentence in (20a), the hearer must depend on inference based on the context, whether overall what is being talked about is something that happened in the past or a current situation. In English, though, as English has grammaticalized obligatory tense marking, the equivalent of (20a) would be (20b), (20c), or (20d), all of which constrain the interpretation of the time frame. (As can be seen from this example, the identification of the gender of the referent (and therefore the identification of the referent) of some pronouns is also constrained by the form of the pronoun, and this too in Chinese is unconstrained.)

(20) a. Tā qù xuéxiào.
        3sg go school
     b. She went to school./He went to school.
     c. She is going to school./He is going to school.
     d. She goes to school./He goes to school.

We can see that compared to Chinese, English obligatorily constrains the interpretation of the time frame, limiting the identification to either a past or non-past situation, but within those broad categories, say, for example given a past tense form, to determine how far in the past the action was the interpreter of the utterance must rely on linguistically unconstrained inference. That is, if I say I ate lunch, then you will probably draw the inference that I ate within the last hour or two, or at least within today; if I say I went to the doctor, then you may make the inference that it was within the last few days; if I say I went to Tibet, then you will not make the inference that it was within the last one or two hours, or even within the last few days, as it could have been quite some time ago. The search for the proper interpretation of the length of time from an overtly marked past action to the time of the speech act is not further constrained grammatically in English. If we then compare English to Rvwang, we can see that in Rvwang there is a four-way past tense system which marks whether the action took place an hour or two ago, a few hours ago but within this day, sometime from yesterday up to a year ago, or more than a year ago. The examples in (21) all are of the verb dī 'to go'.

(21) a. àng dī á:m-ì 'S/he left, went away (within the last 2 hours).'
        3sg go DIR-IP
     b. àng dī dár-ì 'S/he went (within today, but more than two hours ago).'
        3sg go TMhrs-IP
     c. àng dī ap-mì 'S/he went (within the last year).'
        3sg go TMdys-IP
     d. àng dì yàng-ì 'S/he went (some time a year or more ago).'
        3sg go TMyrs-IP

The point is that languages differ quite a lot in how much they constrain the search for the most relevant interpretation, and in what aspects they choose to constrain. As can be seen from these examples, while Rvwang constrains the interpretation of the time frame more than English, it does not constrain the search for the referent of a pronoun as much as English

does. From this we can see that we can not talk about LANGUAGES as being more or less grammaticalized or their interpretation more or less constrained, only particular FUNCTIONAL DOMAINS being more or less grammaticalized or their interpretation more or less constrained in a certain language.

An interesting three-way contrast of what is or is not left to inference in different languages can be seen from a comparison of Chinese, Tagalog and English. The normal way of saying 'Let's go' in Chinese involves just a verb and a particle, as in (22a), and only the use of the hortative particle constrains the interpretation of the actor referent (so it could be 'you go' or 'we go', but not 'he goes'); in Tagalog, as in (22b), it is normal to just say *Tayo na,* which is the 1st person plural inclusive pronoun plus a change of state marker, with no verb, and leave the interpretation of the action suggested unconstrained (it could mean 'Let's go' or 'It's our turn'), while in English both the pronoun and the verb must be specified, so the interpretation of the actor and the action are both constrained.

(22) a. Zǒu bā!
     go   hortative particle
     'Let's go.' or '(Why don't) you go.'

   b. Tayo   na!
      1pl.incl CSM
      'Let's go.' or 'It's our turn.'

Grammatical marking can also function to constrain the interpretation of an utterance by canceling implicatures. For example, the most unmarked interpretation of (23a) would be that the speaker had direct evidence for the claim that the sky had cleared, but this implicature can be canceled by adding extra lexical items, such as the modal verbs *should be/have* or *must be/have*, as in (23b), or a whole clause, as in (23c). In a language with evidential marking that marks assumptions arrived at through deduction differently from those gained through direct experience, such as Qiang, the canceling of the implicature can be done with a single grammatical particle, as can be seen in comparing (24a) and (24b) (from LaPolla 1997a).

(23) a. The sky has cleared.
     b. The sky should have cleared (by now).
     c. The sky has cleared, I guess.

(24) a. mə   tə-χqa-ji.
        sky   DIR-clear-CSM
        Ṭhe sky has cleared.̊(direct experience)

   b. mə   tə-χqa-ji-k.
      sky DIR-clear-CSM-INF
      Ṭhe sky has cleared.̊(guess based on inference rather than direct perception)

Languages can also differ in terms of the type of grammaticalization used to constrain the interpretation of a particular functional domain. For example, given the Chinese sentence in

(25a), the interpretation that the hair being washed belongs to the person doing the washing is purely a matter of inference, as given the right circumstances (such as a professional hair-washer in a barber shop) it could mean the person is washing someone else's hair, but in the English and Rvwang examples in (25b) and (25c) the interpretation is constrained by the obligatory use of the genitive phrase on the object and the reflexive/middle marking respectively.

(25) a. Tā zài   xǐ fǎ
          1sg PROG wash hair
          'S/he is washing (her/his) hair.' (Lit.: 'S/he is washing hair.')

      b. He is washing his hair.

      c. àng nī   zv́lshìē
          àng nī   zv́l-shì-ē
          3sg hair wash-R/M-NP
          'S/he is washing her/his hair.'

In both English and Rvwang the interpretation of the relationship between the actor and the undergoer is constrained, but by very different grammatical categories.

There are a number of other functional domains, such as reflexives and agent-patient differentiation, where English forces certain interpretations and Chinese does not; see LaPolla 1993 for discussion.[3]

4. Implications for the Nature and Genesis of Grammar. We have discussed above that within Relevance Theory certain types of discourse particles, discourse connectives, and performative verbs[4] are said to involve procedural (rather than conceptual) information, in that they guide the hearer in creating the proper context of interpretation. One of the main points I would like to make here is that we can see from arguments made by Gumperz, and

---

[3]In earlier papers (LaPolla 1990, 1993, 1996) I have compared patterns of syntactic behavior in Chinese with those in accusative, ergative, active, and Philippine type languages, and have shown that Chinese does not pattern like any of those systems. I have argued this is because Chinese has not grammaticalized a syntactic pivot for any of its constructions. Chinese therefore should not be considered accusative, ergative, active, or of the Philippine type, but is it another type, possibly called a 'neutral' type, or is it a non-type? Given the facts mentioned above, and others of a similar nature, the tendency has been to see Chinese as another syntactic type, to try to make a syntactic relation out of topic or topic chain (Huang 1989, Shi 1989, Her 1991), or to see 'topic prominence' as a syntactic type in opposition to 'subject prominence' (as many have done based on Li & Thompson's (1976) original proposal of these concepts). I would like to argue instead that a lack of evidence of constraints such as we find in Chinese is precisely that, a lack of constraints. When we say 'type', we mean a set of constraints of a certain type, and if a type is a set of constraints, then the lack of evidence of constraints in Chinese is evidence of the lack of a type, not a separate type. There are ways that Chinese has grammaticalized constraints that English has not, such as numeral classifiers, but in terms of the constraints associated with subject in English, Chinese simply has not grammaticalized them, and so what we get is less constrained inference.

[4]Speech acts are considered a second order act of communication which essentially helps the hearer to process the first order communicative act. The form of the utterance as declarative, imperative, or interrogative alone does not determine the use of the utterance as saying, ordering, or asking, though it can assist in determining the relevance of the utterance (see Wilson & Sperber 1993, Sperber & Wilson 1996).

from the data just presented here, that it is not just discourse particles, connectives and performative verbs that are involved in constraining the search for relevance. Many or possibly all types of grammatical marking are involved in constraining the search for relevance, as they limit the number of possible interpretations that an utterance might have. Grammatical marking (including conventionalized intonation patterns) can have the same effect on the process of interpretation as connectives and lexical phrases in terms of influencing the processing of utterances, as we saw in (5), and in, for example, canceling implicatures, as we saw in (24).

I would also like to argue, again in line with Gumperz, though I came to this conclusion independently, that it is not just the search for implicature that is constrained by grammar, but the proper identification of the communicative act and the explicature as well. As we saw in the discussion above, there are several aspects to interpretation according to Relevance Theory: the recovery of the logical form, which is said to be simply a matter of decoding, the filling out of the logical form to create the explicature, and the creation of implicatures (implicated premises and conclusions). The inferential process is said to start AFTER the recovery of the logical form, because most work done in Relevance Theory assumes the Chomskyan view of linguistic competence as an autonomous module, and the logical form is the output of that module.[5] I would like to argue that in fact there is no 'logical form', as envisioned by those holding to an autonomous module view of syntax; there is simply the recovery of the phonetic or visual form, which is then enhanced into the explicature.[6] We can see evidence of this fact from experiments where a sound or even a word in a sentence was replaced with a cough or other sound, and hearers could still interpret the utterance properly (e.g. Warren 1970). They must then be using inference to fill in the gaps. This is also involved in the need of non-fluent speakers of a language for clearer and louder speech, as they do not have the knowledge of the language and culture to be able to fill in the gaps in the utterance that are created by ambient noise. In terms of the explicature, inference is needed to determine the proper referents, the proper time frame, the proper spatial frame, etc., and all of these inferences can be constrained by the use of either more explicit lexical items, by intonation, or by grammatical elements that constrain the search for the proper interpretation, such as we saw in the examples presented above.

Givón (1979a, Ch. 5; 1979c) has argued that language develops from pragmatic, loosely structured linguistic modes to tighter, more structured modes, and that these modes can be seen in the differences between child language and adult language, between pidgin languages and standard languages, between spoken and written registers, between informal and formal registers, and between unplanned and planned discourse. Communication in the pragmatic

---

[5]See for example Kempson's (1988b:19) statement that 'The theory with which relevance theory shares most assumptions is the Chomskyan paradigm'; see also Espinal 1996.

[6]Sperber and Wilson's (1986/1996:185ff) own discussion of the relationship between the linguistic input module and the central inferential mechanisms in the identification of the correct semantic representation (logical form) could be used as an argument against the modular view. Recognizing that the modular view they accept (that of Fodor 1983) requires that a module be informationally encapsulated, and therefore would not have access to general encyclopedic information or the non-domain-specific processes, such as inference, but at the same time recognizing that the identification of the correct semantic representation must involve contextual information, including general encyclopedic information and inference, they suggest that the input module and the central inferential mechanisms interact on a constituent by constituent basis. It would be much simpler to abandon the modular view of grammar in favor of one that views a speaker's knowledge of grammar as constructed using the same central inferential mechanisms as those used in interpretation.

mode depends largely on word order and the lexicon alone, while communication in the more syntacticized mode depends much more on conventionalized constructions and morphology in tightly structured relationships. Grammar develops as the originally free collocations of lexical items become fixed and (in some cases) lexical items are reduced in form and content until they become morphological rather than lexical forms. Paul Hopper (1987, 1988) developed this idea into the conception of grammar as 'emergent' from discourse. He argued that rather than taking grammar as a given (what he called 'a-priori grammar'), and then possibly seeing how discourse can affect grammar after it is established, linguists should see discourse as prior to grammar, and giving rise to grammar, as repeated patterns of discourse develop into what we think of as grammar. Grammar is then not seen as fixed structure, but something that is constantly evolving (see also Langacker 1987, Ono & Thompson 1995).

What I would like to argue is that the discourse patterns that lead to the development of grammar are those that are repeatedly used for constraining the interpretation of utterances in a particular way. For example, in Old English the word *lic* 'like' was used so often after an adjective to make explicit an adverbial relation to a verb that it became conventionalized and developed into the adverb forming suffix *-ly* used obligatorily in many contexts in English today. In Mandarin Chinese the frequent use of a preverbal locative phrase where there was an implicature of an on-going event led to the development of a progressive marker from the locative verb *zài*. What begins as a conversational implicature over time becomes conventionalized, so it is then a conventional implicature, and then can become further conventionalized until it is simply a part of the grammar that forces a particular interpretation. A somewhat different example, in that it involves an extension of the use of already existing morphology rather than the creation of new morphology, is the extension of the use of the reflexive marker to middle situations such as we saw in the Rvwang example in (25c). A marker that originally was used only to mark reflexives came to be used in some middle situations optionally with an emphatic sense to narrow the range of possible interpretations, and later came to be used so often that it became obligatory for many verbs (LaPolla 1995b), much as what happened in the Romance languages.[7] What we think of as a grammatical construction (or 'constructional schema'—Langacker 1987; Ono & Thompson 1995) is also simply a pattern of usage that was 'instantiated frequently enough that it [now] has a cognitive status independent of any particular context' (Ono & Thompson 1995:219).[8]

The fixing of repeated patterns into grammar is nothing more than the development of conventionalized forms that restrict interpretation, and Givón's cline of forms from more pragmatically to less pragmatically based types correlates with the degree to which interpretation is constrained grammatically rather than lexically. Some of these

---

[7]Once this happens, there is then no formal distinction between reflexives and middles, and so some languages then reinforce or renew the direct reflexive marking, again being driven by the desire to constrain the interpretation. This has happened, for example, in Dutch (Kemmer 1993).

[8]As with so many other things, Bolinger (1961) was ahead of his time when he argued for something like schemata, what he called 'idioms', and combinations of schemata, what he called 'syntactic blends' to form new syntactic structures, and attempted to show 'the permeation of the entire grammatical structure by threads of idiom' (p. 366). He argued against a purely generative view of grammar, suggesting that our use of grammar was partly creative and partly a matter of memory: 'At present we have no way of telling the extent to which a sentence like *I went home* is a result of invention, and the extent to which it is a result of repetition, countless speakers before us having already said it and transmitted it to us in toto. Is grammar something where speakers 'produce' (i.e. originate) constructions, or where they 'reach for' them, from a preestablished inventory, when the occasion presents itself? ... Probably grammar is both of these things ... ' (p. 381).

developments function to constrain the interpretation of the explicature, such as gender distinctions in pronouns, syntactic subjects, tense marking, etc., while others function to constrain the interpretation of implicatures, e.g. discourse connectives, conditionals, concessives, etc., that is, particles and other morphology that help in the determination of what has been called 'speaker meaning' as opposed to 'sentence meaning'.

This brings us to a fourth line of inquiry in linguistics that is relevant to this topic, and that is the work on grammaticalization by Elizabeth Traugott (Traugott 1982, 1988, 1989, 1990; Traugott & König 1991). Independent of the work in Relevance Theory and the work by Gumperz, Traugott has shown that there is a type of secondary grammaticalization where a form that has grammaticalized from a lexical item and at first has only an objective meaning later further grammaticalizes in the direction of subjective (speaker-oriented) meaning, with a stage in between of marking textual cohesion (that is, the path of development is 'propositional ((> textual) > (expressive))'; Traugott 1990:497).[9] The order of change is unidirectional. In Traugott 1989, three semantic-pragmatic principles of grammaticalization are given:

Semantic-pragmatic Tendency I:
Meanings based in the external described situation > meanings based in the internal (evaluative/perceptual/cognitive) situation. (p. 34)

[e.g. *behind* (body part) > (space) > (time), where it operated twice]

Semantic-pragmatic Tendency II:
Meanings based in the described external or internal situation > meanings based in the textual situation. (p. 35)

[cohesive, e.g. *æfter* 'following behind' first became a temporal connective (tendency 1), then became a marker of textual cohesion as a subordinator]

Semantic-pragmatic Tendency III:
Meanings tend to become increasingly situated in the speaker's subjective belief-state/attitude toward the situation. (p. 35)

[e.g. English *sippan* 'after, from the time that', through conversational inference from the temporal sequence came to have a causative meaning (*since*: *Since you've got a cold, we'll cancel the trip*); also *while*: OE *pa hwile pe* 'at the time that' > ME *while (that)* 'during' > PDE *while* 'although'; situation viewed as existing in the world > signal of cohesive time relation between two clauses > expression of speaker's attitude (Traugott 1990:497)]

All three types of grammaticalization involve greater specification of information, movement 'in the direction of explicit coding of relevance and informativeness that earlier was only covertly implied' (Traugott & König 1991: 212). In the first two types of

---

[9]The terms 'propositional', 'textual' and 'expressive' are based to some extent on Halliday and Hassan's (1976; see also Halliday 1985/94) 'experiential', 'textual' and 'interpersonal' metafunctions of language, respectively. Traugott's view of the development of subjective/expressive meanings has certain parallels in Halliday's conception of 'grammatical metaphor' (Halliday 1994, Ch 10).

grammaticalization, you have relatively concrete concepts being used as models for specifying more abstract concepts, and so they are examples of metaphor, but in the last type, Traugott & König argue that what you have in the development of causals, concessives, and preference or denial connectives is the strengthening of informativeness and the conventionalizing of conversational inferences. This strengthening of informativeness is said to be a type of metonymy, where metonymy is said to be 'indexical', i.e., it points to semantic relations in certain contexts, though in this case the contexts are pragmatic contexts of conversational and conventional inference, and '[t]he "indexing" involved is the pointing to the relevance that conversational inferences about stereotypical situations entail.' (p. 211)

Though Traugott and König's view is one of polysemy rather than monosemy, and they discuss subjectification from the point of view of the speaker expressing her perspective on what is said rather than the speaker constraining the interpretation of the hearer, what Traugott & König are talking about here as 'explicit coding of relevance and informativeness' is parallel to what in Relevance Theory is talked about as constraints on relevance, with the prime example being discourse connectives of the type Traugott has been talking about as developing from the metonymic specification of meaning. And while some of the wording is somewhat different, for example using 'stereotypical situations' rather than 'frames', the conception of 'indexing' by grammatical markers presented by Traugott & König is parallel with Gumperz's discussion of contextualization cues signaling the contextual inferences to be used in interpretation.

The movement from objective to subjective meaning is an attempt on the part of the speaker to more and more explicitly express her perspective on what is said, and in doing this she influences the interpretation of the utterance by more and more specifically constraining the possible interpretations. That is, objective lexical items which simply have connotative meaning less specifically influence interpretation than the subjective forms do. The subjective forms have developed regular procedural meanings. Traugott's outline of the development of grammaticalized forms from more objective to more subjective meaning (Tendency I -III) corresponds to the fixing of constraints on the interpretation of the explicature (Tendencies I-II) vs. the fixing of constraints on the creation of implicatures (Tendency III), as laid out in Table 1.[10]

_____

[10] I do not have a separate level of Logical Form in Table 1, as the concept of Logical Form is due in Relevance Theory to the acceptance of the modular view of grammar. As I am assuming that the entire communicative process involves inference, and the entire process can be constrained by grammaticalization, there is no level of 'recovery of Logical Form'.

Level of interpretation	Type of grammaticalization
Explicature	e.g. gender distinction in pronouns, subject, tense marking; helps in the interpretation of sentence meaning˚ (Tendencies I-II)
Implicature	e.g. discourse particles, contextualization cues; helps in the interpretation of speaker meaning˚(Tendency III)

Table 1: Correlations between level of interpretation and type of grammaticalization

I would like to suggest that what Gumperz, Traugott, and Sperber and Wilson are arguing for is basically the same thing, that non-conceptual aspects of language serve to constrain the interpretation of communicative acts. What is different among these three is that Sperber and Wilson assume the Chomskyan view of grammar being an autonomous module, and inference starts after the recovery of the Logical Form, while for Gumperz inference is involved at all levels of the communicative activity. Gumperz has also shown that it is not just a small set of items such as clause-connectives that are involved in constraining interpretation. Traugott has demonstrated the unidirectional development of languages toward greater and greater constraining of interpretation, and has shown the specific path by which some of these 'procedural' meanings develop.

I would argue for a synthesis of the lines of inquiry that I have presented above. That is, we should accept the emergent grammar view of the development of language structure, and ground it in a theory of ostensive-inferential communication such as that provided by Relevance Theory. The development of grammar out of repeated discourse patterns and the development from objective to subjective meaning then can be seen as the fixing of and honing of (as they involve further specification) constraints on the search for relevance during the process of interpretation.[11] The development is not teleological, any more than the evolutionary development of species is. It is in fact a type of evolution, though an aspect of socio-cultural evolution rather than biological evolution.

It has been said that languages differ not so much in what they can say, but in what they must say. This is looking at it from the speaker's point of view. From the hearer's point of view, we can say that languages differ not so much in what can be understood, but what must be understood.[12] All languages can constrain interpretation, but some languages have

---

[11]Though I am presenting this from the point of view of constraints on interpretation, I do not assume that linguistic change is hearer-driven. From one point of view we can say it is speaker-driven, as the patterns can only become conventionalized if speakers choose to use the patterns over and over again. From another point of view the conventionalization process takes time, and involves the same people as speakers and hearers. That is, a speaker uses a particular pattern and other people pick up on that (we are creatures of habit and imitation), and repeated use of that pattern by a number of people causes it to become grammaticalized (such as the same-subject interpretation of English clause-coordination).

[12]Some have also talked about this in terms of responsibility for communication being successful: In a language like German or Russian, where there is a relatively great amount of obligatory grammatical information marked in each clause, the speaker has to do relatively more work in constructing an utterance than in a language like Chinese. On the other hand, in a language like Chinese, where interpretation depends largely on inference, the

developed obligatory grammatical rather than non-obligatory lexical means for doing so, and these grammatical means, in the case of morphology, have developed out of repeated use of lexical means. The function of both lexical and grammatical means in constraining interpretation is the same. For example, the expression *I guess* add to an English declarative clause such as *I guess he's coming* might be said to be adding conceptual information, while adding an evidential particle marking a guess to a similar clause in some other language that has grammaticalized evidential marking would be considered as adding procedural information, yet the function/information of both is the same. Lexical items would not grammaticalize into so-called procedural information if their conceptual information was what was important. The conceptual/procedural distinction then is not a simple either/or situation, but more of a privative opposition (with gradations): all lexical items can have a procedural function, while grammatical markers generally only have a procedural function. That is, some elements involve only so-called procedural information, but lexical items are not used exclusively for conceptual information; they often have a procedural function, and it is having this function that allows them to become grammaticalized into items with only procedural information.

The development of particular grammatical forms is not obligatory or teleological, but it is no coincidence that, for example, the Qiang people of Sichuan, China, live on the sides of mountains along river valleys and their language (Qiang; Tibeto-Burman) has developed a complex system of direction prefixes including prefixes marking 'up-river' vs. 'down-river' and 'up the mountain' vs. 'down the mountain', with the latter two being the two most frequently used direction prefixes (LaPolla 1997a). There are also sometimes competing motivations (DuBois 1985) for one pattern or another, but the process of a particular form becoming conventionalized is the same. For example, English lost the distinction between singular and plural second person pronouns because of a repeated pattern of using the plural pronoun when referring to a singular referent out of politeness considerations, but some Southern (U.S.) dialects have regrammaticalized a second person plural form *y'all* from repeated use of *all* after *you* to clarify when *you* was referring to a plural referent.

Christian Lehmann has argued that '[w]e must reverse our basic perspective and regard the linguistic system not as given, but as created by language activity.' (1985:314). According to Lehmann, language activity is goal directed, and therefore systematic ('it is, in fact, a constant systemization'—p. 314), but it is also creative, as new ways of achieving the goals of communication are found. He has attempted to explain why there are regular patterns and stages of grammaticalization, and why grammaticalization is unidirectional, as being due to the fact that on the one hand speakers want to be more creative and expressive, but on the other hand are subject to the constraints of the linguistic tradition of their communities. That is, there is a set of conventions that speakers of a language must to some extent conform to, but speakers also desire to be original in their expressions. He explains what he calls the 'channelization of grammaticalization' (1985:315), the fact that generally the same lexical items cross-linguistically are grammaticalized into the same types of grammatical marking, e.g. demonstratives into definite articles, and the numeral 'one' into indefinite articles, as being due to 'requirements of semantic aptitude imposed on elements which are to be grammaticalized' (315). The result is that speakers will tend to use the same

---

listener must do comparatively more work in determining the intended interpretation. This is why some scholars (e.g. Huang 1984, LaPolla 1995c) have talked of Chinese being a 'cool' language (like a cool medium) and German as a 'hot' language.

devices to enrich their utterances because of the limited number of devices available for use. (See also Lehmann 1990, 1992 on the relationship between particular grammaticalization processes and social factors.)

The view I am presenting here is very much in line with that of Lehmann, except that I would argue that aside from the need to say things more expressively and vividly, there is a need to say things clearly, and the 'requirements of semantic aptitude' are precisely the lexical meanings needed to make the expression more explicit, and thereby constrain interpretation, at the stage of using lexical items to constrain interpretation, and then the repeated use of these lexical items causes them to become grammaticalized.

What I am arguing for involves a different way of viewing structure. Rather than assuming that communication is unambiguous and treating ambiguous expressions as aberrant, we should assume that forms used in communication are inherently ambiguous (Reddy 1979), and look at structure from the point of view of how it constrains interpretation, that is, how it is made less ambiguous by, for example, the grammaticalization of subject or other grammatical categories. Most linguistic studies, even many of those that use natural language data rather than made-up sentences, still take the grammar as given, and only look for the 'interface' between semantics and syntax or pragmatics and syntax. For many, such as Susumo Kuno (e.g. 1987) and Ellen Prince (e.g. 1988), pragmatics is simply another module of the package, and not the foundation of communication and therefore of grammar.[13] The view I am presenting here is that the fundamental aspect of communication is not the linguistic structure, but the interaction of the speaker and hearer in performing a communicative activity. The role of the context in the performance of this activity involving the interpretation of utterances is not to simply supplement semantic meaning; the context is the base on which all communicative activity depends. Communication can occur without linguistic structure; linguistic structure is simply an instrument used to help the interpreter infer the speaker's indended meaning.[14]

To take one example of what I mean by looking at grammar in a different way, we can look at Ekkard König's (1995) study of the meaning of converb constructions. This is an excellent paper, but it focuses not on how the constructions constrain the interpretation of an utterance, but instead focuses on how the constructions are vague and so need to be enriched by contextual factors. That is, König took the form of the converb construction as something basic and then tried to see how contextual factors help us to interpret the meaning of the converb construction, in a sense treating it as if its grammatical form was simply another concept-bearing lexical item, and says that general background assumptions and contextual information and general principles of language use 'make an important contribution to an interpretive enrichment of the nonspecific basic meaning of converbs.' (p. 83). Instead we can look at the utterance and try to interpret the speaker's communicative intention, and see how the use of a particular structure, such as a converb construction, constrains our search for the proper interpretation, that is, how the use of a particular grammatical form constrains our search through general background assumptions and contextual information and general principles of language use in order to help us find the intended interpretation.

---

[13]For arguments against the modular view of pragmatics, see Wilson & Sperber 1986.

[14]Georgia Green has said that "communication is not accomplished by the exchange of symbolic expressions. Communication is, rather, the successful interpretation by an addressee of a speaker's intent in performing a linguistic act" (1996: 1). This is not just true of linguistic acts, but of all communicative acts.

The view of grammar I am presenting here means not trying to define what, for example, a 'subject' is, the way Keenan (1976) did, but seeing what we call 'subject' in English as a set of constraints on the interpretation of certain syntactic constructions such as clause coordination, etc., and seeing which constraints individual languages have or have not grammaticalized as part of their grammatical system (see Van Valin & LaPolla 1997, Ch. 6). It also means not inventing covert movements and structures to try to explain all differences of interpretation as differences in syntactic structure.

This makes explanations of not only why a particular type of marking develops possible, but also why the use of marking that has already developed becomes extended in predictable ways, such as the development of agentive marking from ablative marking or the extension of reflexive marking to middle situations (see LaPolla 1995b). The development is in the direction of greater specificity and a more constrained set of possible interpretations, utilizing resources already present in the language when possible.

This view represents a synthesis of a number of independent lines of inquiry in linguistics. Independent lines of research in a field can only converge in this way when there is an empirical basis for the conclusions reached, and the convergence shows we are making solid progress in our understanding of the object of inquiry. One line of research that does not fall together with the ones mentioned above is the non-empirical Chomskyan view of a 'Universal Grammar' which is a genetically endowed set of structure types (parameters).[15] The view I am presenting here negates the Chomskyan view, as it sees grammar not as genetically endowed, but as a socially evolved set of conventions. Saying that constraints on interpretation such as those we associate with subject in English, or for a language to have a set word order (so-called 'configurationality'), is due to the setting of genetically determined parameters makes no more sense than saying that restrictions on the eating of pork among Semitic people, or the wearing of hats by Semitic men, are due to genetically determined parameters of pork-eating and hat-wearing.[16]

5. Summary and Conclusions. Communication, linguistic or otherwise, is based on inference. A speaker (communicator) performs an ostensive act in order to communicate. This gets the attention of the hearer (interpreter), and alerts him/her to the fact that the speaker has a communicative intention. The hearer then must work to first recover the (visual or acoustic) signal that the speaker has produced, and then attempt to make sense of it, first arriving at what has been called the 'explicature', then determining the implicatures which must be created in order for the explicature to make sense in the context of the communicative activity. All of this is done using inference; all aspects of interpretation involve the creation of a set of assumptions, a context, which can be added to whatever part of the signal or message has been recovered up to that point (it is a dynamic process) to deduce the most likely form and interpretation. Communication is not a simple matter of coding and decoding. The inference involved in communication is essentially guesses at

---

[15]See Comrie 1981/1989, Ch. 1 for discussion of the non-empirical nature of Chomsky's notion of 'Universal Grammar'.

[16]As I understand it, at some point in the distant past, before Jews and Arabs went their own ways, there was the idea that one should always wear a hat when one says the name of God. As people wanted to be safe, they started wearing hats all the time, and this 'culturalized' (became a mandatory part of the culture), so that the rule is now that people should wear hats all the time. The non-eating of pork 'culturalized' in a similar way: originally there was a functional motivation (avoidance of illness), but over time it simply became obligatory.

what the communicator's intended message might be. These guesses are based on the principle of relevance, i.e. the assumption that an ostensive act involves a guarantee of relevance, and that the communicator will chose the form for the ostensive act that will most likely lead to the intended interpretation. The speaker then will tailor the utterance, in the case of linguistic communication, in such a way that the hearer will not have to expend unnecessary effort to create a context that will allow him/her to arrive at the intended interpretation. In doing this, the speaker takes into consideration guesses as to what information is available to the hearer at the time of utterance for use in interpreting the utterance. The most straightforward reflection of the latter aspect is the amount of lexical content that the communicator includes in the utterance. Grammaticalized marking (including intonation) can also be used to help the hearer process the conceptual information given by the lexical items by constraining the search for relevance. Though grammaticalized marking differs from the lexical items in not being mainly for expressing conceptual content, the grammatical marking performs the same role in constraining or guiding the interpretation of the utterance that an increase in the number of lexical items can have.

Grammar develops from repeated patterns of language use, and is often extended in its use by metaphoric and metonymic processes. Grammatical marking develops out of repeated use of lexical items that eventually get fixed in a particular structure, and then begin to lose their phonetic and lexical integrity. A second aspect of grammar is the development of constraints on structures such as those we associate with the concept of 'subject' in English, for example the same-subject constraint in clause coordination with a reduced second clause. Languages begin without such constraints and without any grammatical marking, so at that stage interpretation relies purely on the use of lexical items and pragmatically determined word order. With time repeated discourse patterns get fixed in the language, leading to the grammaticalization of grammatical marking or the fixing of particular structural constraints. At this stage interpretation is more constrained; still involving inference, but inference constrained by the grammar. While the development is not teleological, it can be said that structure develops to constrain interpretation, as the initial repeated pattern that eventually became fixed as grammar was used to constrain the interpretation. As grammar is then an evolved set of social/cultural conventions (part of the larger set of social conventions we think of as language), languages come to differ in terms of what will be constrained and what will not, and in terms of the degree to which interpretation of a particular functional domain is constrained.[17]

We should then not take grammar as basic and try to interpret grammar, but take inference as the basis of communication, and try to determine how grammar develops to constrain interpretation.

I have argued that within Relevance Theory the basic idea of constraints on the search for relevance should be expanded to cover all aspects of grammar, and that Relevance Theory

---

[17]Following are some of the implications of this view for linguistic theory:

1. As languages differ in terms of the ways they constrain interpretation, both in terms of constraining or not constraining a particular type of interpretation (functional domain), and also in the degree to which the interpretation is constrained and how it is constrained, the differences between languages are gradient differences, not simple parameters;

2. As these constraints are the result of grammaticalization, they are therefore not genetically hard-wired;

3. (a second order conclusion) The human language ability then can not be an autonomous and genetically programmed module; language development and use must be based on general cognitive structures.

should embrace a functionalist view of language structure and development. I have tried to show that the different lines of inquiry discussed here can be seen as complementing each other, often talking about different aspects of the same phenomena, and so can be unified within a single comprehensive view of communication, cognition and language development.[18]

References

Blakemore, Diane. 1987. Semantic constraints on relevance. Oxford: Basil Blackwell.

Blakemore, Diane. 1988a. 'So' as a constraint on relevance. In Kempson, ed. 1988a, 183-195.

Blakemore, Diane. 1988b. The organization of discourse. In Newmeyer, ed., 1988, Vol. II, 229-250.

Blakemore, Diane. 1990. Constraints on interpretations. Proceedings of the 16th Annual Meeting of the Berkeley Linguistics Society, 363-370.

Bolinger, Dwight L. 1961. Syntactic blends and other matters. Language 37:366-381.

Comrie, Bernard. 1981/1989. Language Universals and Linguistic Typology (second edition, 1989). Oxford: Blackwell, and Chicago: University of Chicago Press.

Comrie, Bernard. 1988. Coreference and conjunction reduction in grammar and discourse. Explaining language universals, ed. by John Hawkins, 186-208. Basil Blackwell.

Dixon, R. M. W. 1980. The Languages of Australia. Cambridge, London, and New York: Cambridge University Press.

Du Bois, John W. 1985. Competing motivations. In Haiman, ed. 1985, 343-365.

Fodor, J. A. 1983. The modularity of mind. Cambridge, Mass.: MIT Press.

Givón, Talmy. 1979a. On understanding grammar. New York, San Francisco, London: Academic Press.

Givón, Talmy, ed. 1979b. Syntax and Semantics, Vol. 12: Discourse and Syntax. New York: Academic Press.

Givón, Talmy. 1979c. From discourse to syntax: grammar as a processing strategy. In Givón, ed. 1979b, 81-112.

Green, Georgia. 1996. Pragmatics and natural language understanding, second edition. Mahwah, NJ: Lawrence Erlbaum Associates.

Gumperz, John J. 1977. Sociocultural knowledge in conversational inference. Linguistics and anthropology, Georgetown University Round Table on Languages and Linguistics 1977, ed. by Muriel Saville-Troike, 191-211. Washington, D.C.: Georgetown University Press.

Gumperz, John J. 1982. Discourse strategies. Cambridge: Cambridge University Press.

Gumperz, John J. 1989. Contextualization cues and metapragmatics: The retrieval of cultural knowledge. CLS 25, pt. 2: Parasession on Language in Context.

Gumperz, John J. 1992a. Contextualization and understanding. Rethinking context, ed. by Alessandro Duranti and Charles Goodwin, 229-252. Cambridge: Cambridge University Press.

Gumperz, John J. 1992b. Contextualization Revisited. The contextualization of language, ed. by Peter Auer & Aldo di Luzio, 39-53. Amsterdam and Philadelphia: John Benjamins Pub. Co.

Gumperz, John J., Gurider Aulakh, & Hanna Kaltman. 1982. Thematic structure and progression in discourse. Language and social identity, ed. by J. J. Gumperz, 22-56. Cambridge: Cambridge University Press.

Espinal, M. Teresa. 1996. On the semantic content of lexical items within linguistic theory. Linguistics 34:109-131.

---

[18]Steve Nicolle (to appear) also argues for looking at grammaticalization using the point of view of Relevance Theory, but the approach is somewhat different from the one presented here. Nicolleaccepts the modular view of the mind and its implied decoding model of the initial stage of interpretation, and argues for a strict differentiation between conceptual and procedural encoding, with the latter constraining the interpretation of the former. What I am arguing for here is a completely inferentially based view of communication where both so-called conceptual and procedural information constrain the interpretation of the speaker's intention.

Haiman, John, ed. 1985. Iconicity in syntax (Typological studies in language 6). Amsterdam and Philadelphia: John Benjamins Pub. Co.

Halliday, M.A.K. 1994. An introduction to Functional Grammar, second edition. London & New York: Edward Arnold.

Halliday, M.A.K. & R. Hasan. Cohesion in English. London: Longman.

Her, One-Soon. 1991. Topic as a grammatical function in Chinese. Lingua 84:1-23.

Hopper, Paul. 1987. Emergent grammar. Proceedings of the 13th Annual Meeting of the Berkeley Linguistics Society, 139-55.

Hopper, Paul. 1988. Emergent grammar and the A Priori Grammar Postulate. Linguistics in context: Connecting observation and understanding, ed. by Deborah Tannen, 117-134. Norwood, NJ: Ablex.

Hopper, Paul J. & Sandra A. Thompson. 1984. The discourse basis for lexical categories in universal grammar. Language 60.4:703-752.

Hopper, Paul J. & Sandra A. Thompson. 1985. The iconicity of 'noun' and 'verb'. In Haiman, ed. 1985, 151-183.

Huang, C.-R. 1989. Subcategorized topics in Mandarin Chinese. Paper presented at the 1989 CLTA Annual Meeting, Nov. 17-19, Boston, MA.

Huang, C.-T. James. 1984. On the distribution and reference of empty pronouns. Linguistic Inquiry 15.4:531-574.

Keenan, Edward L. 1976. Towards a universal definition of 'subject'. In C. N. Li, ed. 1976, 305-333. New York: Academic Press.

Kemmer, Suzanne. 1993. The middle voice (Typological studies in language 23). Amsterdam and Philadelphia: John Benjamins Pub. Co.

Kempson, R. M, ed. 1988a. Mental representations: The interface between language and reality. Cambridge: Cambridge University Press.

Kempson, R. M. 1988b. The relation between language, mind, and reality. In Kempson ed., 1988a, 1-25.

König, Ekkehard. 1995. The meaning of converb constructions. Converbs in cross-linguistic perspective, ed. by M. Haspelmath & E. König, 57-95. Berlin & N. Y.: Mouton de Gruyter.

Kuno, Susumu. 1987. Functional syntax: anaphora, discourse, and empathy. Chicago: University of Chicago Press.

Langacker, Ronald W. 1987. Foundations of Cognitive Grammar, Vol. I: Theoretical prerequisites. Stanford: Stanford University Press.

LaPolla, Randy J. 1988. 'Subject' and referent tracking: Arguments for a discourse-based grammar of Chinese. Proceedings of the West Coast Conference on Linguistics (Fresno, October 14-16, 1988), Volume I, ed. by Joseph Emonds, P.J. Mistry, Vida Samiian & Linda Thornburg, 160-173. Dept. of Linguistics, California State University, Fresno.

LaPolla, Randy J. 1990. Grammatical relations in Chinese: Synchronic and diachronic considerations. PhD dissertation, University of California, Berkeley.

LaPolla, Randy J. 1993. Arguments against 'subject' and 'direct object' as viable concepts in Chinese. Bulletin of the Institute of History and Philology 63.4:759-813.

LaPolla, Randy J. 1995a. Pragmatic relations and word order in Chinese. Word order in discourse, ed. by Pamela Downing & Michael Noonan. Amsterdam and Philadelphia: John Benjamins Pub. Co.

LaPolla, Randy J. 1995b. On the utility of the concepts of markedness and prototypes in understanding the development of morphological systems. Bulletin of the Institute of History and Philology, Academia Sinica 66.4:1149-1185.

LaPolla, Randy J. 1995c. Review of Understanding Utterances: An Introduction to Pragmatics, by Diane Blakemore, Oxford, UK & Cambridge, USA: Blackwell, 1992. Language 71.1:173-176.

LaPolla, Randy J. 1996. On the place of Chinese in the typology of grammatical relations systems. Paper presented to the Linguistics Seminar of the Dept. of Chinese, Translation, and Linguistics, Oct. 31, 1996.

Lehmann, Christian. 1985. Grammaticalization: synchronic variation and diachronic change. Lingua e Stile 20:303-318.

Lehmann, Christian. 1990. Grammaticalization and related changes in contemporary German. Approaches to grammaticalization, ed. by Bernd Heine & Elizabeth C. Traugott, Vol. 2, 493-535. Amsterdam & Philadelphia: Benjamins.

Lehmann, Christian. 1992. Word order change by grammaticalization. Internal and external factors in syntactic change, ed. by Marinel Gerritsen & Dieter Stein, 395-416. Berlin: Mouton de Gruyter.

Li, Charles N. & Sandra A. Thompson. 1976. Subject and topic: A new typology of language. Subject and topic, ed. by C. N. Li, 459-489. New York: Academic Press.

Mishra, A. 1980. Discovering connections. Proceedings of the 6th Annual Meeting of the Berkeley Linguistics Society, 73-89.

Newmeyer, F. 1988. Linguistics: The Cambridge survey. Cambridge: Cambridge University Press.

Nicolle, Steve. 1997. Conceptual and procedural encoding: criteria for the identification of linguistically encoded procedural information. Proceedings of the University of Hertfordshire Relevance Theory Workshop, 27-29 Oct. 1995, ed. by Marjolein Groefsema, 47-56. Chelmsford: Peter Thomas & Assoc.

Ono, Tsuyoshi & Sandra A. Thomspon. 1995. What can conversation tell us about syntax? Alternative lingusitcs, ed. by Philip W. Davis. Amsterdam & Philadelphia: Benjamins.

Prince, Ellen F. 1988. Discourse analysis: A part of the study of linguistic competence. In Newmeyer, ed. 1988, Vol. II, 164-182.

Reddy, M. J. 1979. The conduit metaphor—a case of frame conflict in our language about language. Metaphor and thought, ed. by A. Ortony, 284-324. Cambridge: Cambridge University Press.

Selting, Margret. 1992. Intonation as a contextualizing device: Case studies on the role of prosody, especially intonation, in contextualizing story telling in conversation. The contextualization of language, ed. by Peter Auer & Aldo di Luzio, 233-258. Amsterdam & Philadelphia: Benjamins.

Shi, Dingxu. 1989. Topic chain as a syntactic category in Chinese. Journal of Chinese Linguistics 17.2:223-262.

Sperber, Dan & Deirdre Wilson. 1986/1996. Relevance: Communication and cognition. (second edition, 1996) Cambridge, Massachusetts: Harvard University Press.

Traugott, Elizabeth C. 1982. From propositional to textual and expressive meanings: Some semantic-pragmatic aspects of grammaticalization. Perspectives on historical linguistics, ed. by W. P. Lehmann & Y. Malkiel, 245-271. Amsterdam & Philadelphia: Benjamins.

Traugott, Elizabeth C. 1988. Pragmatic strengthening and grammaticalization. Proceedings of the 14th Annual Meeting of the Berkeley Linguistics Society, 406-416.

Traugott, Elizabeth C. 1989. On the rise of epistemic meanings in English: An example of subjectification in semantic change. Language 65:31-55.

Traugott, Elizabeth C. 1990. From more to less situated in language: the unidirectionality of semantic change. Papers from the 5th Int. Conf. on English Linguistics (CILT 65), ed. by S. Adamson, V. Law, N. Vincent, & S. Wright, 497-517. Amsterdam & Philadelphia: Benjamins.

Traugott, Elizabeth C. & Ekkehard König. 1991. The semantics-pragmatics of grammaticalization revisited. Approaches to grammaticalization, vol. 1, ed. by E. Traugott & B. Heine. Amsterdam & Philadelphia: Benjamins.

Uhmann, Susanne. 1992. Contextualizing relevance: On some forms and functions of speech rate changes in everyday conversation. The contextualization of language, ed. by Peter Auer & Aldo di Luzio, 297-336. Amsterdam & Philadelphia: Benjamins.

Van Valin, Robert D., Jr. & Randy J. LaPolla. 1997. Syntax: Structure, meaning, and function. Cambridge: Cambridge University Press.

Warren, R. 1970. Perceptual restoration of missing speech sounds. Science 167:392-393.

Wilson, Deirdre & Dan Sperber. 1986. Pragmatics and modularity. Proceedings of the 22nd Regional Meeting of the Chicago Linguistics Society, Pt. II, 67-84.

Wilson, Deirdre & Dan Sperber. 1993. Linguistic form and relevance. Lingua 90:1-25.