

## 9 The Learner's Problem of Arranging Words

*Wolfgang Klein and Clive Perdue*

"I see another language as distorted English and then I try to work it out." (attributed to a linguist)

In the course of the acquisition of a second language in everyday communication, the learner passes through a series of more or less elaborate repertoires of linguistic devices that allow him to express himself and to understand others with varying degrees of success when he tries to communicate with his social environment.<sup>1</sup> Repertoires of this sort we call *learner varieties*. We may assume that both the internal organization of a learner variety and the transition from one variety to the next are systematic in nature and that because of this systematicity, both "the internal organization of learner varieties" and "the logic of development" are problems worth studying.

This paper deals with the first of these problems, the internal organization of learner varieties, and more particularly with the question of how learners arrange words in their utterances if their repertoire is still very limited and most of the normal syntactic devices of the target language are not yet available to them. Its aim is therefore to attempt to find out what other organizational principles they can call on to make themselves understood. We think that the investigation of learner varieties should eventually deepen our understanding of the acquisition process, through a study of the second problem: the logic of development. Moreover, it should also cast some light on language and its functioning in general, and it is this latter aspect that motivates the present study.

Why should the investigation of the use of almost pathologically restricted systems such as elementary learner varieties tell us something that could not much more easily be uncovered by looking at normal, fully developed language, whose investigation is more advanced in many respects? In full-fledged languages, the interplay of forms and functions is extremely intricate and tight - they have a much higher degree of integration when compared to

The study reported here is part of a larger crosslinguistic project on adult second language acquisition. First analyses of the data were presented at a colloquium "Acquisition d'une langue étrangère" in Aix-en-Provence, May 1984, and at the University of California, Los Angeles, February 1985. We are grateful to our colleagues within the project, and to the participants of the Los Angeles seminar for their help and advice. We would especially like to thank Maya Hickmann, Eric Kellerman, Brian MacWhinney, and Christiane von Stutterheim for their detailed comments.

learner varieties, in which there are only a limited number of lexical items, and where the means available to indicate their interrelation within more complex units are dramatically restricted. The advantage of learner varieties is that it is relatively easier to disentangle the web of forms and functions and to study their interplay.<sup>2</sup>

This becomes particularly clear when looking at the way learners arrange words in their utterances. It has often been argued that for any type of language use descriptions in purely syntactic terms - case marking, government, or word class membership - do not suffice to cover all of the regularities that govern the internal organization of utterances in discourse. It has also been argued that nonsyntactic organizational principles - such as "information distribution" or "theme-rheme structure" - emerge particularly clearly in learner varieties simply because syntactic devices expressing these functions are less available: A learner variety which has no inflection cannot use case marking or agreement, for example.

But how can we develop a sound analysis of this aspect of utterance organization, given that concepts such as "theme," "focus," or "background" are much less clear and solid than, for example, morphological case marking? The obvious thing to do is to give them a more precise definition, or even better, to adopt or to develop a conceptual framework of which they constitute an integral part, and where they are related to syntactic concepts, such as "subject," "verb," "nominative," and so forth. It would then be possible to apply concepts of both types to the analysis of data from specific languages or varieties and to try to uncover the language-specific principles which determine the make-up of utterances in that language or variety. By comparing the language-specific givens, we may eventually be able to state more general constraints on the internal organization of utterances in discourse.

There is no a priori reason why such a procedure should not work. But in practice, it has led to an impressive mess. Over the last ten years, there has been much discussion about topic-oriented versus subject-oriented languages (kindled, especially, by Li & Thompson, 1976), and the outcome of this discussion is not fully convincing. It seems that despite considerable efforts to define these terms and related ones, different authors analyze the same phenomena in different ways; and their application across languages is anything but consistent and therefore lacking in comparability. Maybe this need not be so, but it *is* so. Consider, for example, the notion "subject": In languages such as English,

We do not claim that learner variety utterances are in any sense closer to some "underlying semantic representation." If it is difficult - as we shall see — to substantiate claims about the overt structure, one can hardly see how we could make claims about the structure of the underlying representation; as a consequence, statements about "closeness" between these two levels ("higher semantic transparency") seem hard to verify. (Personally, we agree with Lancelot & Arnault, 1664, that written French best mirrors one's thoughts.)

German, or Italian, for which a long research tradition exists, there is usually little disagreement over what the subject of a given sentence is (for simple clauses at least); but by and large, the subject is case marked and movable in German, not case marked, movable and occasionally "null" in Italian, and neither case marked nor movable in English. In what sense, then, are we dealing with the same phenomenon? (For a careful examination of the relevance of "subject" for German, see Reis, 1982.) Needless to say, the situation is much less clear for "topic," where even for the most studied languages it is often debatable what the topic of a given utterance is. Consequently, we feel that universal statements based on "topichood" or "subjecthood" should be treated with suspicion.

The situation is even worse if we try to analyze learner variety utterances using such categories directly. What criteria allow us to label a lexical item "subject" rather than "topic" in an utterance which lacks all morphological marking, whose word order could be otherwise in a particular variety, and which even lacks a finite verb? Constructions of this sort are quite typical for learner varieties, as we shall see in what follows. Note that the problem is not just to establish a clear definition of these concepts which makes them consistently applicable to all languages; it is equally important to have a workable operationalization which allows for consistent analysis of given utterances in individual languages (and varieties), especially if these languages have received little or no description so far, as is clearly the case with learner varieties. Thus, von Stechow (1981) gives a clear and convincing definition of "topic" and "focus" in terms of formal semantics (see also Klein & von Stechow, 1982); but there is no direct way from these definitions to the concrete determination of what the "topic" in a given utterance is. We do not wish to say that the present unsatisfactory situation can never be remedied; we are simply saying that the many attempts along the lines mentioned above have led to numerous practical problems.

In what follows, we will try a different, more modest approach to the problem of arranging words in utterances. It is basically inductive and has at least the advantage of being controllable and easily applicable. In the next section, we will briefly explain what we have in mind.

### **The Learner's Problem of Arranging Words**

At any point in time, an adult learner (as any other speaker) can draw on different kinds of cognitive resources whenever he wants to communicate. First, the adult learner already knows a language. This allows him to draw upon both the specific expressive devices which constitute that language and the semantic and cognitive categories that underlie them, such as modality, deixis, agency, or whatever. Second, the adult learner already knows about language and communication in general. He knows that he has to monitor for communicative success, that speaker and hearer have differing background knowledge, that there

are social conventions for determining who is allowed to talk to whom, and so forth. Possibly, he does not know how these functions are concretely realized in the language he has to learn and in the culture which he enters (or invades), but he knows that such rules and influences do exist. Third, the adult learner has a lot of nonlinguistic information about the world and, thanks to his eyes and ears, about the situation in which he is communicating. Fourth, he knows bits and pieces of the target language from what exposure to it he has had. Obviously, these resources may be very different from learner to learner, and also change for any given learner over time.

Suppose now that you are an adult Italian worker who has been living and working in England for about six months. You have already learned, perhaps, a number of proper names, such as *John*, *Peter*, and *Mary*, as well as some base forms of verbs, such as *love*, *come*, and *kill*. You have learned that the proper nouns denote human entities and that the verbs denote the actions these entities can accomplish. You have no inflection, hence no case morphology or agreement (that is, your learner variety is quite restricted). Suppose now that on some occasion and for some reason, you want to express the thought that Peter is in love with Mary. You are able to denote some components of this thought, Mary by *Mary*, Peter by *Peter*, and that relation by *love*. But this does not suffice: somehow, the words must be arranged. This follows from the fact that language is linear (there are no hierarchical complications here). But this arrangement must also be done in such a way that the listener is able to derive the intended thought from what is uttered.

There are two extreme possibilities with respect to the six theoretically possible arrangements. One possibility is that there might be no restrictions at all in your learner variety. In this case, it is unlikely that you will get your message across since your utterance is in many ways ambiguous. In this case, the language analyst has nothing to say with respect to the organization of your utterance, he must just wait for a more advanced learner variety. The second possibility is that there may be restrictions of some sort. The obvious question then is: In what terms can they be stated? This is the case that will interest us.

We and others have considered a variety of possible influences on the selection of the first noun in our hypothetical learner variety utterance. (Sridhar, this volume, discusses a number of such influences on sentence formulation in adult native speakers.) Among the most important are:

- I. a. The shortest unit (in terms of phonemes) is first. Behaghel (1923) proposed this as a general influence on all of syntax. However, we take this tendency to be a consequence of other facts.
- b. The verb comes first. This case makes sense only if the learner indeed distinguishes verbs from other word classes in his variety and if the analyst can perceive the distinction. This is by no means trivial", we might ask what a "verb" is if there is no tense, inflection, or agreement. Thus, one might argue

- that the distinction between *love* (verb) and *love* (noun) totally collapses in this variety (Klein 1984).
- c. That NP which is morphologically unmarked (nominative) comes first This is actually impossible in this particular variety, as it was defined above, since it has no morphological marking at this time. So, if "subject" is defined as a syntactic category on morphological grounds, then there is no point in speaking of subject in this variety at all.
  - II. a. If the thought to be expressed corresponds to an action or activity, then that entity which performs the action ("agent") is named first. If there is no action or activity, the choice is free.
    - b. Animate entities are named first; if there are more than one of them, the choice is free, unless one of them is human; then this one comes first.
  - III. a. An entity which was referred to before comes first; if there are more than one of them, the choice is free.
    - b. That entity which first comes to the speaker's mind comes first.
    - c. That entity which is dearest to the speaker's heart comes first.
    - d. That entity which the speaker thinks to be best known to the listener comes first.
  - IV. Some combination of these influences determines the word order. For example, if the thought in the learner's mind involves an action with an animate and an inanimate participant, then the animate participant comes first unless the inanimate participant was mentioned before, then this one comes first, and so on.

The last possibility is envisaged by Bates and MacWhinney (this volume) when they note that fully developed languages integrate several of these influences into particular "coalitions." They hold that concepts such as "subject" and "topic" can be defined on the basis of such combinations. Learner varieties can also express such coalitions, as we shall see.

The actual impact on learner varieties of the kinds of influences I—III listed above is open to speculation. The alternative approach, mentioned above, to an *a priori* definition of, for example, "topic" would try to determine which influences are indeed operative in utterances in discourse, and then aggregate them to complex interactions of these influences, exemplified by IV. It is here that the analytic task is easier for the study of learner varieties than of full-fledged languages. As a competent communicator, the adult learner will use the limited possibilities he has to maximal effect, and the principles underlying this use will emerge more clearly precisely because the linguistic means they apply to are limited.<sup>3</sup> Thus, a careful analysis of learner variety texts should provide both a deeper insight into the inner organization of learner varieties and an exemplification of the general principles that obtain in the organization of adult native languages. The definition of concepts such as "subject," "theme,"

One can imagine a mirror argument for child language acquisition (see, e.g., Karmiloff-Smith, 1981): The child uses a relatively greater command of vocabulary and utterance-internal structures to solve the relatively more complex problem of arranging the utterances into coherent discourse, for example, by manipulating individual utterances in order to allow a "discourse theme" to appear consistently in utterance-initial position.

and "focus" should then derive from the regularities that could be observed in studying these varieties.

What causes a learner to obey one of these influences in his learner variety production rather than another? Various answers are possible; we will mention and briefly comment upon four of them. The first answer to this question is that it could be the case that the influence already holds in the learner's first language and that this influence is simply transferred over to the target language. This is probably the most common explanation, and there is no doubt that transfer of this sort can influence the structure of learner varieties. However, for transfer to occur, the learner must both perceive (however wrongly) a possible L1-L2 equivalence and have *some L2 means* to operationalize this perception in production. In this sense, transfer resembles many other domains in that, the more you know, the more kinds of mistakes you are able to make.

A second possible reason for adopting one of the influences noted above is that the learner assumes that it holds in the target variety. This sounds almost trivial. Why, after all, should a learner use a rule if he does not think it holds in the language to be learned? But in real-time communication, the learner is often forced to apply rules he is totally uncertain about or which he even thinks to be false with respect to the target language. This situation is familiar to any speaker of a second language; but it has also been reported for child language (see Klein, 1983, Chapter 8).

A third possible reason for obeying a particular influence is that the influence may be based on a universal constraint. Claims to this effect have often been advanced in the literature (for example, Gass, 1984; Rutherford, 1984); they are sometimes seen as an alternative to transfer hypotheses (although the choice is not mutually exclusive). There are various ways in which we might conceive universal constraints to operate.

1. Constraints could function as "generative" universals in accord with a Chomskyan "universal grammar." The proposals here involve concepts such as "subjacency," "specified subject condition," and so forth. (See Chomsky, 1982; Hornstein & Lightfoot, 1981; Felix, 1984). If there are indeed constraints of this sort, then they do not say very much about utterance organization in learner varieties that are as elementary as those discussed here. It is hard to see how a constraint such as "subjacency" or even a universal phrase structure constraint could restrict<sup>4</sup> the possible word orders in the "Peter is in love with Mary" example.
2. Constraints could function as "statistical" universals, in the sense of the Keenan-Comrie hierarchy of noun phrase accessibility (Keenan & Comrie, 1977; Hawkins, 1983; Gass, 1984; Comrie, 1981). Universals of this kind may serve as heuristic guidelines: They give the researcher an idea of where to look for interesting

Note that this is not an argument against the existence of such constraints or even the constraining force of universal grammar in language acquisition, both first and second. Universal grammar in this sense resembles a husband who comes to the kitchen and offers his helping hand when the washing up is done except for three spoons and a saucepan.

phenomena. But they raise some of the same problems as UG (e.g., not all learner varieties actually have relative clauses) and all of the problems which we briefly discussed above. So long as we do not have criteria for what a "subject" is in learner variety utterances (and elsewhere), universals of this kind are of little help (Perdue, 1984).

3. Constraints could be "pragmatic" universals of the type enumerated in III, such as "from known to unknown" (Behaghel, 1923-32) or "me first" (Cooper & Ross, 1975; see also Silverstein, 1976b). We think that universal influences of this sort indeed play an important role in the organization of learner varieties. They are not particularly clear, however, and we think it might be more practicable to describe first what is indeed operative in learner varieties in these terms and then look for possible extensions and generalizations of the results, rather than stating a universal and then applying it to learner varieties. Thus the study of learner varieties may help in providing a better description of these universals, as we suggested above.

For both the generative and statistical universal constraints, the same objections hold that have been made in connection with transfer and its possibility. To be operative, both kinds of universals require that the learner have considerable knowledge of the language to be learned. This is much less the case for pragmatic universals which are less tied to the specifics of a given language and as such can be assumed to be broadly shared by competent communicators. These pragmatic constraints therefore tie in well with a fourth, and rather different, answer to the question why a learner might prefer a specific constraint over another, which is that the learner may assume that following this constraint makes his utterance better understandable (or understandable at all). This possible and plausible answer is unlike the first three in that it speaks of the speaker's state of knowledge of the listener. Having determined the listener's state of knowledge, the speaker may still need to rely on what he knows about pragmatic and statistical constraints.

We think that, in reality, all of these factors influence the way in which the learner organizes his utterances, and the way in which they interact with each other can only be determined by comparing learner varieties under varying conditions: with different target languages, different source languages, and at different developmental stages. The following study is a step in this direction.

Before turning to the data in detail, it might be useful to have a more global look at the procedure. In what follows, we shall analyze three relatively long texts from three learners, all of them foreign workers who developed their learner variety by everyday contact in the host country rather than in the classroom. The L1-L2 combinations are Italian-German, Italian-English, and Spanish-French; that is, there are three target and two source languages. The text is a retelling of a part of a silent movie to another person, who had seen the preceding part of the movie together with the learner. This defines a controllable background - we have at least a partial control of the speaker's and the listener's shared knowledge. We also know the general shape of the information that the speaker is trying to communicate. Obviously, our understanding of this information is

not perfect, but it is reasonably clear. We have chosen to analyze a full text, rather than individual sentences, because a full text allows us to control for introduction, maintenance and shift of referents under varying circumstances. We also have a reasonable control over the temporal and spatial organization of the whole story. This kind of data gives us good, although by no means ideal, material for verifying a whole series of possible influences or constraints of the types listed above. There are three methodological problems, however, which deserve mention:

1. The data analyzed here are limited both in size and type. It may well be that the learner, under different circumstances, organizes his utterances in a different (or partly different) way. This can only be determined by examining other types of data. Indeed, in the context of the project in which this study was done, a wide range of data was collected; we plan to extend the present analysis by including other text types at a later stage.
2. It is often very problematic to interpret utterances in learner varieties. There is no "native speaker" that could be asked whether a certain construction is impossible, or (and this is far more problematic) what a certain utterance that was recorded some time ago really means. As a consequence, most samples of learner varieties contain certain utterances that are wildly ambiguous or simply uninterpretable.
3. Finally, we are faced with all the practical problems of analyzing recorded spoken language. Learner variety samples show many false starts, hesitations, or self corrections. This is not a problem in principle (and indeed may provide us with helpful additional insights), but in practice, it is a challenge for any reliable analysis. We will be confronted with numerous problems of this sort. In general, the examples presented in this paper are "edited" to exclude obvious false starts, breakdowns, hesitations, interjections, and metalinguistic questions such as "correctly speaking?" Readers who wish to study the full un-edited corpus are welcome to contact us directly.

So much for the aims and the general background of the present study. Now we will move from the fog of theoretical considerations into the swamp of empirical facts.

### **The Data**

Our data come from the European Science Foundation project entitled "Second Language Acquisition by Adult Immigrants." The project was set up as a coordinated, comparative study taking place with identical schedules and identical data-collection procedures in five European countries, England, France, Germany, the Netherlands, and Sweden. In each country, we studied the acquisition of the local language by adult speakers of Arabic, Spanish, Finnish, Italian, Punjabi and Turkish. The pairings of source languages (SL) and target languages (TL) are such that comparisons can be obtained on the acquisition of one TL by speakers of two different SLs as well as the acquisition of two different TLs by speakers of the same SL: We can have, therefore, at least some means of distinguishing in the acquisition process between phenomena specific to one SL-TL configuration and more general phenomena. The overall aim of

the project is to isolate the factors which determine the structural properties and tempo of the acquisition process in four major domains of investigation, one of which is the arrangement of words in learners' utterances. For a complete description of the aims and methodology of the ESF project, see Perdue (1984b).

The present paper reports on a small part of the full database collected in the larger project. Here we present a pilot analysis of one specific activity - a film-retelling task. This activity took place during the first half of data-collection with the three informants mentioned above.

### *The Charlie Chaplin Study*

The project researchers working in Heidelberg made a montage of extracts from Charlie Chaplin's film *Modern Times* which lasts about twenty minutes and is divided into two main episodes, described below. The procedure for the activity is simply that a researcher and an informant watch the first episode together, then the researcher leaves the room while the informant watches the second episode. The researcher returns immediately after the end of the film in order to listen to the informant retelling the second episode.

*Episode 1: America 1930 - Poverty, Hunger, and Unemployment.* Charlie gets into a demonstration against unemployment, is mistaken for the leader and put into prison. At dinner one of his fellow prisoners hides heroin in the salt cellar, and Charlie helps himself to it by mistake. The drug gives him a heroic force: He foils an attempt to escape and frees the director, who, in gratitude, releases him with a letter of recommendation for a job. Charlie is not too enthusiastic about this because he feels he is better off in prison than at liberty. Parallel with this we see a second story: A young girl (whose father is a widower, unemployed, and without the means to feed his three children) steals food for her family. Her father is shot in a demonstration, and the children are sent to an orphanage. The girl manages to escape at the last moment.

*Episode 2: Determined to Return to Prison.* Charlie finds work in a shipyard. He clumsily causes the launching of a ship that was not finished. He is immediately fired and is all the more determined to go back to prison. The girl roams through the streets, hungry, and steals a loaf of bread. When she tries to escape she runs into Charlie and both fall to the ground. A woman, who watched the theft, calls the baker. A policeman comes to arrest the girl. Charlie tries to claim responsibility for the theft but it doesn't work. The girl is marched off to prison. Charlie tries again to get back into prison. He goes into a restaurant, eats as much as he can, calls a policeman from the street and tells him that he has no money to pay the bill. He is arrested. In the police car he

again meets the girl who stole the bread. In an accident they are both thrown out of the car. The girl suggests that he escape with her, and he does. They rest for a while in the garden of a middle-class house, and watch the couple who live there say a tender good-bye to each other in front of their house. Charlie and the girl dream of such an existence. A few days later the girl has a surprise for Charlie: she has found a house. Of course, it is a ruined cabin in a miserable condition, so that a series of hilarious accidents happen when they first come to see it. But they don't let this disturb their happiness. In the last scene we see them walking down a long road that disappears into the horizon.

If we assume that the learner participates cooperatively, that is, that he wants to try to get his listener to understand the story he retells rather than confuse him, then this task is interesting for our present purposes for two major reasons. First, this is a complex verbal task: The speaker retells part of a relatively complex story, consisting of events whose relationship to each other must be specified. Within each event, the speaker has to tell who did what to whom, introducing new characters and maintaining reference to characters who are already on stage. The main characters are male (Charlie) and female (the girl) and they act and are acted upon. Their stories, which run in parallel during the first half of the film, intertwine during the second half, necessitating a choice on the part of the learner as to which of them is central in each event.

Second, this task is interesting because, as in the study by Sridhar (this volume), we have partial control over what is mutually known or unknown to the learner and the listener at the beginning, in that we have the film to compare with his production. This partial control gives us, therefore, a partial idea of what the speaker *maximally* wants to retell (given the cooperation assumption above). The comparison between an imposed message and the speaker's actual production is highly interesting as it is safe to assume that he does not have the TL means to fulfill his wish (or what we hypothesize to be his wish). On a conceptual level, it is then at least plausible that he will reorganize the story to a certain extent in order to achieve the double (and not necessarily compatible) purpose of accommodating his message to the means available while ensuring comprehension on the part of the listener. It is in this interplay of conceptualizing and formulating that the semantic/pragmatic "universals" we listed earlier may emerge particularly clearly in his speech, since both are available to him to a large extent irrespective of his knowledge of the TL, and assumed by him to be available to his listener. Thus, at the inter-utterance level, discourse organization principles (von Stutterheim, 1986) carry much communicative weight, and interact with utterance-organizational principles such as those mentioned earlier, because it is these principles that are the least dependent on specific elements of a TL.

### Italian-German: Vito

After some remarks on Vito's biography and communicative behavior, we will briefly sketch his linguistic repertoire at the time of the interview and then give a detailed analysis of approximately his first twenty-five utterances describing the "shipyard" episode in order to illustrate both the typical features of this learner variety and the practical problems involved in this type of analysis.

Vito was born in 1948 near Palermo (Sicily). At the time of the recording (1983), he had been living in Germany for about eighteen months, but still had a very limited command of German, which reflects his very limited contacts with German speakers. His wife is Italian, they have no children, and he works in the kitchen of an Italian restaurant. He is however talkative, outgoing, and very interested in things linguistic, as his metalinguistic behavior shows. While retelling the film, he often interrupts himself and asks for a word or expression with the "formulaic" questions: *Was ist der Name?* "What is the name," *Was Name diese?* "What name this?" Occasionally, he checks the correction of his own speech: *Richtig spreche?* "Correctly speak?" We also find less apparent traces of metalinguistic activity: (1) his production seems carefully planned, with clear prosodic phrasing of each word; (2) when quoting speech - a metalinguistic device he and other informants often use - his production appears to be closer to the German standard than when he is reporting events or providing background material. This variation leads us to distinguish three types of utterances: those based on formulaic metalinguistic speech such as *Was ist die Name?*, those involving the narration proper including the story line and background material, and those based on quoted speech. The distinction between these three types of utterances is best illustrated by Vito's use of the copula: It never occurs in narrative utterances, it sometimes occurs in quoted speech, and it frequently occurs in formulaic utterances as part of the formula. In what follows, we will omit formulaic utterances from analysis as Vito does not himself "arrange" the words in them, and point to other differences between narrative utterances and quoted speech. Some rare instances of quoted speech are apparently induced by the intertitles of the film. The intertitles were only briefly shown on the screen and written in three languages (German, Turkish, and Italian); one would assume Vito to focus on the Italian version, but on occasion he clearly registered and reproduced the German one.

### *Vito's Linguistic Repertoire*

Considerations of space force us to be very brief here. In particular, problems encountered in assigning words to classes are not dealt with; some of them emerge clearly in the next section. For a detailed analysis of Vito's lexicon see Dietrich (1984) and for an exhaustive analysis of Vito's means of referring to space see Becker (1984). Some overall aspects of Vito's system are:

1. Vito has no inflexional morphology (except in some instances of quoted speech), hence no case marking, no agreement, and no tense.
2. In this text, he uses about sixty different nouns, forty verbs, a dozen adjectives (in epithet and attribute function), and about ten adverbs of time, place and modality.
3. He uses three articles - *da* "the", *diese* "this, that", and *eine* "a" - and, very rarely, the quantifiers *viel* "much", *all* "all", and *zwei* "two". *Eine* is also used as a numeral.
4. He has a minimal pronominal system (Klein & Rieck, 1982): *ich* "I", *du* "you", *mir* "me" (after a preposition, and possibly *Sie*, polite "you", these appearing essentially in quoted speech. Otherwise he has just one pronominal form (*sie*) and one adjectival form (*seine*) for the third person. *Zusammen* "together" is occasionally used as an argument of the verb to denote Chaplin and the girl. *Diese* can be used alone, deictically and anaphorically, under conditions described in the next section.
5. He uses one preposition - *in* - very frequently, and half a dozen others occasionally. *In* is highly overgeneralized to denote all kinds of spatial relations. His negatives are *nix*, which is frequent, and *keine*, which is less so. The latter, a TL determiner, is used by Vito before nouns, but also as an alternative to *nix* as a sentence negator. Both may be used together in one utterance.
6. Finally, he uses the utterance connectors *und* "and", *aber* or *\*pero\** "but", and *oder* "or". To mark a restart, he uses *\*alora\** "then", which is the only Italian word he uses with any frequency. We ignore these connectors when assigning words to positions (initial, final) within the utterances examined below. (In what follows, source language sequences are bounded by \*, pauses indicated by +, and a short text omission indicated by ...).

### The Shipyard Episode

The retelling task starts with the scene where Chaplin leaves prison with a letter of recommendation. Chaplin and the letter are both mentioned in the interviewer utterance immediately preceding V1:

V1.     *sie habe brief + brief für gefängnis*  
           she have letter     letter for prison

The intended meaning of the utterance is fairly clear - Chaplin has/had a/the letter from prison. Vito's lack of tense-marking makes it impossible to determine the tense of the verb in this passage. For simplicity's sake, we will therefore use present tense forms in the glosses. The utterance structure before the pause is NP1 V NP2 :

1. NP1 refers to Chaplin. *Sie* corresponds to a pronoun of the TL ("she" or "they") although the appropriate TL-form would be *er* "he". As Chaplin was mentioned in the immediately preceding utterance, Vito starts his task with an anaphoric pronoun (in preverbal position).
2. V denotes a stative relation.
3. NP2 also denotes a previously mentioned entity, but the internal structure of this NP is N. We cannot however draw any firm conclusions as to the referential status (definite or indefinite) of this N: It would be equally appropriate to refer back to - "the letter" - or reintroduce - "a letter" - in this context.

NP2 is related to the (appositional) sequence after the pause. It is hard to see, at this point, whether this sequence is a mere "postscript" or rather a disguised

relative - "...letter, (which was) from..." Note, however, that *gefängnis* is a bare noun and is situationally defined (both interviewer and informant know about the prison from watching the first half of the film together). As a shorthand, we will sometimes use the term "thematic" to indicate a referent that has been explicitly referred to, pointed to, or which any adult speaker would infer, to the exclusion of other possible referents, in a given situation.

V2. *komme in eine baustell + baustell vielleicht*  
 come in a building site building site perhaps

The intended meaning before the pause is again clear: Chaplin comes to a building site. The structure is V PP, where PP is directional. *Komme* denotes an action and seems to have standard German meaning here; we infer that this was performed by the individual referred to in initial position in the immediately preceding utterance. Note that this does not allow us to state that V2 contains a zero anaphor. What we have are two possible conditions for leaving a referent unexpressed: (1) either it was thematic immediately before, or (2) it was in initial position in the immediately preceding utterance.

The PP corresponds to the TL pattern, except that *eine* is invariant in Vito's variety, thus we cannot assume the correct TL accusative marking. Note that the building site is introduced into the discourse, appropriately, by a noun accompanied by the indefinite article. The sequence after the pause may be glossed as something like: "a sort of building site" (it is of course a shipyard). We take it to be a "postscript" - one of the possible analyses of V1. Note that the now contextually given part of the sequence - *baustell* - is in first position and is not accompanied by an article.

V3. *diese mache schiff*  
 this make ship

The structure of V3 is clearly NP1 V NP2.

1. NP1 refers to an entity mentioned in the immediately preceding utterance. There, it occurred postverbally and had a different semantic role. Thus, we may say that *diese* goes with "reference maintenance" but also with "position shift" and/or "role shift."
2. *Mache* clearly denotes an action, hence *diese* is an agent, although one would not normally consider a shipyard to be an agent; hence, standard semantic processes that allow us to go from "the shipyard" to "the people at the shipyard" also apply to *diese* here.
3. *Schiff* introduces a new referent. It is unclear whether it is specific or generic or singular or plural. V3 can be glossed either "this one is building a ship" or "this is one of those that builds ships". This example therefore still does not allow us to assign clear referential status to bare nouns.

V4. *kleine schiff mache*  
 small ship make

Although the specific versus generic interpretations are still possible, the latter seems less likely, the more so since the "reality check" of the film shows just

one small ship being built. The structure is clear - NP V - but perplexing. NP is a "patient," the agent is unexpressed, we can infer "shipyard" from conditions (1) and (2) of V2 for leaving a referent unexpressed, and again we see no justification for postulating zero anaphora (where would we place it?). But it is hard to see why the order is NP V rather than V NP, unless the choice between the two is random. If it is not random, several possibilities come to mind:

1. There might be a structural principle at work (roughly: if an utterance contains one NP, it occurs preverbally irrespective of its semantic role), but this seems to be falsified by V2 and by numerous subsequent examples.
2. It may be that different semantic roles of NPs are associated with specific positions, but it is hard to see a great difference between the relation of *schiff* to *mache* in V3 and the relation expressed here.
3. More plausibly, the different orders of V3 and V4 may be based on the discourse organizational principle "go from given to new," except that here we have a (focussed?) case of going from "new" to "given" (cf. also the occurrences of *baustell* in V2). But standard assumptions about "given-new" distribution are perhaps too general in any case.

We will forego further discussion of V4 for the present, and return to it later when we can compare it to other similar utterances (V21).

V5.     *chef    arbeiter    rufe: " "*  
           chief-   worker    calls

Here *rufe* introduces the quoted speech of V6, and the quotation marks - " " - indicate that *rufe* is a *verbum dicendi*. The intended meaning is obvious: "the foreman calls" and the structure is quite clear. The NP is definitely referring, introduces a new protagonist, but contains no article: What is meant is "the foreman of the shipyard" reference to the shipyard being left implicit. The semantic relation between this and V is agent-action.

V6.     *"Charlie Chaplin + ich brauche eine holz"*  
           "Charlie Chaplin    I    need    a    wood (log)"

It is clear from the intonation that the first NP is a vocative: We will have no more to say about it. The intended meaning of the sequence after the pause is clear and the whole construction is close to the TL pattern, as is often the case when Vito quotes speech. The structure is again NP1 V NP2, NP1 being a deictic pronoun "in the mouth" of the *chefarbeiter*. NP2 is nonreferential and contains, appropriately, the indefinite article.

V7.     *ich brauche eine (keil)*  
           I    need    a    (wedge)

V7 is a repetition of V6, and is separated from it by a metalinguistic passage in which Vito asks, and gets, the German word for "wedge" (*keil* is in brackets to indicate this):

V8.     *sie    nix    verstehn*  
           she no    understand

The most plausible interpretation of V8 is: "he does not understand," with "he" referring to Chaplin. Although "Chaplin" appears in the preceding utterance (if we ignore the repetition), it is a vocative, and contained in quoted speech, neither of which condition obtains in V8. It seems that conditions 1 and 2 of V2 for leaving a referent unexpressed are quite strict.

V9.     *nix komme eine keil,     eine holz, lang, zu lange*  
           no    come   a   wedge, a    log, long, too long

If our conditions for leaving a referent unexpressed are correct, we have two alternative interpretations: *er kommt nicht mit einem Keil* "he doesn't come with a wedge", or *es kommt nicht ein keil* "it is not a wedge that comes". The second interpretation is perhaps more plausible, and since Vito uses *mit* "with" elsewhere, and with a "presentational" verb such as *komme*, it is not implausible that its argument appears postverbally. There are other instances of such presentational constructions in the text. The whole construction consists of two adversative components, roughly: "come not wedge - (but) log." It is interesting to note that the negation *nix* precedes the whole first clause although it only applies to the NP *eine Keil*. *Keil* and *Holz* are accompanied, appropriately, by the indefinite article: The referent of *eine holz* is introduced into the discourse here, whereas *ein keil* remains nonreferential.

V10.    *und sie spreche " "*  
           and she speak

The context makes clear that *sie* refers to the foreman. Assuming we were correct in interpreting *sie* in V8 as "Chaplin," we have to conclude that *sie* may refer to an entity introduced more than one utterance back in the discourse with the intervening material containing a possible referent for *sie*. *Sie* is again preverbal, and *spreche* introduces the quoted speech of V11 and V12,

V11.    *"diese nix"*  
           "this   no (nothing)"

The meaning is clear, the structure too. The initial NP refers to the log mentioned in V9 - that is, more than one utterance back - although its referential status is complicated by the fact that "in the mouth" of the foreman it is deictic. With the *ich* of V6, *Charlie Chaplin* of V6 and *diese* here, we have items of quoted speech entering referential relations with elements of the surrounding text.

V12.    *ich wolle eine ... (keil)*  
           I    want a       (wedge)

Here *keil* is in brackets because Vito had to ask again for the word for "wedge." The meaning and structure of this utterance are clear; our remarks on the second part of V6 apply here too.

V13.    *sie       gucke eine keil*  
           you/she look a   wedge

It is not clear whether the stretch of quoted speech stops at the end of V12

or V13. The intonation patterns of V12 and V13 are very similar: However, this interpretation necessitates analyzing *sie* as the polite address form, which Vito never uses elsewhere. Note, however, that this interpretation would give another example of Vito's production in quoted speech being relatively closer to the TL norm. Or, the meaning is "He (= Chaplin) looks for/sees a wedge." The indefinite article accompanying *keil* is appropriate for either gloss of the verb, and we have another example of preverbal *sie* jumping over an appropriate referent (*sie* in V10) and taking up a previously introduced referent.

V14. *hinten + eine grosse holz*  
 behind (adv) a big log

This and the following utterances describe a complicated piece of business in the film, where Chaplin tries to remove a large wedge maintaining the timbering holding up the ship. When he succeeds in removing the wedge, the timbering collapses and the ship is launched. The structure of this utterance would correspond to the TL but for the absence of a presentative ("there's") between the adverb and the NP (i.e., "behind, *there is* a big log"). This NP is in final position and introduces - with appropriate *eine N* - a new referent. The structure thus shares one characteristic of the "presentative" interpretation of V9, providing indirect evidence for the plausibility of the latter interpretation. Similar examples follow.

V15. *komme diese nix weg*  
 come this no off/away

It is unclear whether the intended meaning is "this does not come away" or "Charlie can't get this away" (German *bekommen* "get"). The latter interpretation would violate the conditions for leaving a referent implicit (see V2), which hitherto have seemed quite strict. The former interpretation would be totally consistent with previous uses of *diese* if its position were preverbal. One could appeal to V9 to explain the order *komme diese*, but this provokes further problems: *Komme* is not a presentative here, and *diese* is, of course, definite.

V16. *seine hand nix habe keine kraft weg diese*  
 his hand no have no strength off this

"His hand was not strong enough to remove it." We may postulate two parts to this complex utterance: NP1 neg V (neg) NP2, and Adv NP3. NP1 is definite; *seine* seems to function like *sie* - it is (part of) a preverbal NP and can refer back several utterances to a NP already in preverbal position (*sie* of V13 in either interpretation of that utterance). V denotes a stative relation, as in V1. NP2 is either TL-like with *keine* as a determiner, or it can be analyzed as having a constituent negator *keine* corresponding to "not". In the second part of the utterance, *weg* functions as a sort of causative verb on the "patient" *diese*. This could explain the order V - *diese*, but example V18 below further complicates

the picture. Note that *diese* in the utterances V15 and V16 refers to the same entity, that those utterances are adjacent, but that the semantic relation between *diese* and the Vs is different.

V17. *sie gucke eine ... (hammer)*  
 she look a (hammer)

Both meaning and structure are clear; the utterance adds nothing to what we already know.

V18. *probiere diese weg*  
 try this off/away

Again the agent is maintained and left implicit. *Diese* refers back, not to the hammer, but to the log. As in the case of *sie*, it can therefore "jump" appropriate referents, although here it has the same semantic relationship to the "verb" *weg* as in V16, and their relative order is reversed. As in the case of V4 and V9, an appeal to structural, semantic and discourse-organizational principles gives unconvincing answers: We have no real explanation at this point.

V19. *\*pero\* sie nix gucke*  
 but she no look

Structure and meaning are again clear. The agent is maintained but not left implicit, perhaps because of the presence of a connective.

V20. *diese holz sicher schiff*  
 this log safe ship

Vito may have learned the (rare) German verb *sichern* "to make safe". But it is more plausible that he has learned to use the (common) adjective *sicher* "safe" as a verb (cf. *weg* in V16) relating two arguments. It would be pointless to postulate a "recategorization" of *sicher* from adjective to verb in a variety where derivational morphology is virtually nonexistent, and zero markings need as much justification in learner varieties as in other varieties. NP1 here is a kind of instrument and refers to an entity which is thematic, and which when last mentioned was in a patient relationship to the "verb". NP2 - *schiff*- most plausibly takes up the ship introduced in V3 — in the nongeneric interpretation of that utterance, and hence is definite.

V21. *diese schiff arbeite neue schiff bau*  
 this ship work(er) new ship build(ing)

Both the meaning and the structure of V21 are unclear. Vito seems to want to explain that, roughly, "This ship was being worked on ... was in the process of being built." *Diese schiff* is reintroduced explicitly. This, and other examples in the text, indicate that the "all-purpose" pronoun *sie* is in fact restricted in its use to refer to animates. We take *diese schiff* and *neue schiff* to be co-referential, the second NP providing a further specification. If we assume the same process to be at work in V4 above, where we had *kleine schiff*, then the specific, singular interpretation of *schiff* in V3 does seem more plausible.

The words *arbeite* and *bau* are (not surprisingly) morphologically undecidable between N and V, but as Vito's utterance patterns have overwhelmingly been (NP) V (NP) so far, we see little reason to postulate an utterance consisting of a string of NPs. As with *weg* and *sicher*, they are verb-like. Now, the NPs they associate with are in a relation of patient, not agent. There is no expressed agent. Vito appears to have a principle - and we will interpret V4 *kleine schiff mache* under this principle - that with verbs that are two-place (or can be two-place - *arbeiten*), and where the structure conveys the semantic relationship agent, patient, action, then the agent can be left implicit - for whatever reason: inferable in V4, unspecified in V21 - and the patient occupies preverbal position. One could call this a "passive": However, Keyser and Roeper (1984) and others point to a relatively wide-spread "ergative" use of action verbs in nonergative languages, such as in *der Laden schloss* "the shop closed", etc. It seems to us more plausible to posit such an "ergative" use in Vito's variety as it takes into account principles which explain other aspects of Vito's variety - word order, semantic role relationships - rather than appealing to "missing" copulas and verbal morphology in interpreting these utterances as "really" passives, or indeed to some otherwise unmotivated "fronting" operation. Speculative though this may be, it yields a consistent picture.

V22. *Zwei holz... sie weg eine + diese schiff weg*  
 two log... she away one + this ship away

V23. *diese schiff weg + sofort meer... kaputt*  
 this ship away straightway sea destroyed

V22 and V23 close the shipyard episode (for reasons of space, we have omitted two utterances). They mean something like: "There are two logs ... he takes one away and he takes the ship away (the ship goes away). The ship goes away and straight into the sea (straightaway, there is the sea) and ... destroyed." The first part of V22 is a "presentational"; here Vito restricts himself to the focal part of the presentational, as the whereabouts of the logs are easily inferred from the preceding discourse (e.g., V20). The second part again contains the "verb" *weg*, associated with an agent (Chaplin, last referred to in V19 where again he was an agent) and a patient - the quantifier *eine* ("one of them"). The last part of V22 is ambiguous, given the principles we have established so far: Either the agent of the previous part of the utterance is left implicit, and the structure is "ergative," or *weg*, like *arbeit*, is either a one- or two-place "verb," and we have a one-place "verb" of locomotion here, similar to the nonpresentational *komme*. In either case, *diese schiff* functions as we now expect a NP containing *diese* to function: Its referent is inanimate. That verbs which may be one- or two-place function like other one-place verbs or like other two-place verbs does not surprise us. In V23, the first constituent is identical with the last constituent of V22, and the same remarks apply. The second constituent is ambiguous, given what we know so far: Either it is a "presentational" - *da war gleich das*

*meer* "straightaway, there was the sea" - and we have a structure akin to V14, or *weg*, like nonpresentational *komme*, takes a directional adverbial - the bare noun *meer* — itself qualified by a temporal adverbial *sofort*. Finally, Vito says *kaputt*, which not only expresses the demise of the boat, but also the end of the episode.

*Two Conclusions and a Lesson.* Two conclusions and a lesson may be drawn from our analysis of the shipyard episode. The lesson is the following: The methodological problems encountered in this type of analysis are severe. Inferring Vito's intended meaning is not always easy. This has consequences on all levels of analysis:

1. We distinguished two levels of discourse where Vito's production showed differences - the narration proper and quoted speech: assigning V13 to one level or the other leads to differing analyses, as we saw.
2. The segmentation of the speech stream into utterances is not always easy. In V1 we have hesitated between a "relative" and a "postscript" interpretation of the sequence after the pause - *brief für gefängnis*. Suppose now that Vito does indeed show traces of incipient subordination relations: we might then be justified in noting that of the three occurrences of *gucke* in the text, two have two overt NP arguments and one - in V19 - only one, and in assuming that the whole utterance V20 serves as its second argument "he does not see (that) this log supports the ship." But we have as yet no overt indications that Vito's production does contain subordination relationships.
3. Establishing tentative regularities and applying them to further utterances yields alternative interpretations - compare the "intransitive" versus the "ergative" interpretation of *diese schiff weg* in V22.
4. Finally, assigning words to classes (cf. *sicher*, *weg*), and, of course, establishing the meaning of individual words, is sometimes problematic.

These problems are compounded by what we may call the "closeness fallacy" - a methodological trap which creates a tendency to analyze learner variety utterances as minimally deviant from TL utterances. We are not convinced that we have avoided the trap (or indeed, if it is entirely avoidable), but we have come closer to avoiding it than in many other studies (including earlier versions of this paper), where the typical trap we fell into was the following: There is a learner utterance NP1 V NP2, we imagine a "corresponding" TL utterance NP1' V NP2' and note that NP1' is its subject and NP2' its object, *therefore* NP1 is the subject, and NP2 the object, of the learner utterance. What we have *tried* to do here is rather to capture regularities that are present in the text: (a) identify verbs and their overt arguments; (b) characterize the lexical propensities of the verbs and the semantic relationship holding between them and their arguments; (c) characterize the referential status of the arguments; (d) try to find relationships between (b) and (c) on the one hand, and the internal structure and position of the arguments on the other hand. We have not relied on TL grammatical functions such as "subject" or "object", because they are in any case not foolproof, and call on phenomena such as verb agreement and case

marking which are absent from this text, and because we are not tempted by arguments such as: "The subject is in preverbal position (in the TL), therefore the NP in preverbal position is the subject." This has also prevented us, so far, from postulating entities such as "zero anaphora." However, we dare to be less prudent in the following paragraphs.

The first, tentative, conclusion is then that the way Vito forms his utterances is related to the lexical properties of the verb.

1. The verb denotes an action A first subclass - *mache, gucke, bau* - has either one or two overt NP arguments. In the latter case the preverbal NP denotes the "agent" and the postverbal NP the "patient" or "experiencer." If there is one overt argument, then if the patient is missing (one of the interpretations of V19), the rest of the structure remains constant; if the agent is missing, then either the agent is that of the immediately preceding utterance, and the rest of the structure remains constant, or the "ergative" structure occurs, and patient is preverbal. We will return to the ergative construction in a minute in order to see *why* it occurs. A second subclass has one or two overt arguments. These are inherently one-place verbs - for example, *komme* in V2 - and have the configuration agent-V̄, where the agent can be missing under the same conditions as for the first subclass, and where the whole configuration may be accompanied by a postverbal directional. There has only been one clear case so far - V2, there were competing analyses for V15 - but the remainder of the retelling contains many such examples.
2. The verb denotes a stative relationship (*habe, brauche, wolle, sicher*). Here, the structure is always NP1 V NP2, but the semantic relationships entertained between the verb and its arguments are very heterogeneous - possessor-possessed, instrument-patient, and so forth. Overall, one could imagine an "affectedness" hierarchy here, with the more "affected" NP being postverbal. We will return to this later.
3. Finally, there are a number of presentative or equative constructions in which there is not always an overt verb. A presentative consists of the NP that is presented, which may be preceded by an adverbial but need not be, as in V22. We have seen no equational constructions yet: They consist of two arguments, the first denoting what is defined or characterized - always an NP - the second being either an adjective or another NP.

The second conclusion is that of the central constituents of the utterances - NPs and Vs. Vs are invariant, but the internal structure of NPs varies. We may distinguish four cases:

1. NPs consisting of a noun, which is sometimes, but not always overtly determined.
2. NPs consisting of *sie* or *diese*. The former always denotes a human entity and always appears preverbally. The latter denotes inanimate entities and occurs preverbally and postverbally. There is a case we have not encountered yet, where, in postverbal position, *diese* denotes a human entity.
3. In quoted speech, there are some occurrences of *ich*, possibly *sie*, and, further on in the text, *du*.
4. Finally, there are cases where in order to understand the utterance, we have to infer a referent that is not overtly expressed.

We will now attempt to put the two conclusions together. Let us take the case where the two argument "action-verb" construction NP1 V NP2 introduces a participant into the discourse in NP1 position (say, *chef arbeiter*). He is the

agent of the verb therefore, and is thematic. If, at more than one utterance's remove, he again becomes thematic *as agent*, he is referred to as *sie*, and *sie* is of course in preverbal position. The only cases with this construction where we do not get the configuration (NP1, *sie*) V NP2 are:

1. When the agent *remains thematic from the immediately preceding utterance*. We are now in a position to posit zero anaphora, since a structure 0 V NP2 aligns with all other instances of the two-argument action-verb structure, and we can predict under what discourse conditions the first place will be realized as NP, *sie*, or zero.
2. When the *patient* of the preceding utterance becomes thematic, and the agent is irrelevant (that the patient is thematic is marked both by preverbal position and by further determination or qualification). In this case we would *not* wish to posit a "zero anaphora" for the agent: there is not necessarily an appropriate referent available, and the possible configurations containing zero are nowhere matched by other, overt configurations in the text

We may now turn back to the stative verbs, which also enter the configuration NP1 V NP2. Out of context, some of these utterances would seem to be reversible - here, we are appealing to our intuitive knowledge of the world - in the sense that *ich wolle eine keil* and *eine keil wolle ich*, and *sie habe eine brief* and *eine brief habe sie* would convey the same "message" to any (German-speaking) adult. On the other hand *diese schiff sicher holz* is less readily interpretable as "this log holds the ship up." Whatever the "semantic role hierarchy" may be, we conclude that for the less reversible cases, the NP higher up the hierarchy will be in NP1 position, and for the more reversible cases, discourse constraints - for example, "me first," what is thematic at that time - will determine the relative order of arguments.

The preceding two paragraphs may be seen as a first attempt to formulate, inductively, the "coalition" of constraints which govern the arrangement of words in Vito's utterances: For a structure NP1 V NP2, if the verb denotes an action, then NP1 will be filled by the agent, realized as a lexical NP, a pronoun, or zero, and so on. Obviously, these conclusions are tentative, and already have possible counter-examples. But the approach has been illustrated, and provides a good springboard for the more general analysis that follows.

### *Rudolfo and Ramon*

Before turning to the utterance organization of all informants, we will sketch the social background and linguistic repertoire of Rudolfo and Ramon, the other two informants. Rudolfo is an Italian in his mid-twenties. After his "maturita," or high school diploma, in Italy, he worked for some time in an accordion factory, and then went to London, where he had been for 15 months at the time of the interview. After some months' work in an Italian restaurant, he found a job in a coffee-house where the language spoken is English. This job is his main contact with English, others being sport with English friends and one term of courses at a college of further education.

As was the case with Vito, it is necessary to treat separately (a) narration proper, (b) quoted speech, and (c) metalinguistic comments. In what follows, we will only consider (a) and (b), excluding, however, questions and commands (which only occur in quoted speech). Rudolfo uses about 20 verbs, including the copula. They occur in the base form or as base + *ing*; there is one "seen", one "fell", perhaps one "banged"; the copula almost never combines with base + *ing*. He increases his verb repertoire with "onomatopoeia," accompanied by gestures, which we will indicate as < crash>, < whoosh>, and so forth (see Ru13 below). Of the about 150 occurrences of lexical NPs in the text, about half are introduced by "the"; their usage seems to correspond to standard English. Bare nouns (except the name "Charlie") refer typically to indefinite or generic uncountables ("ham", "bread", "work"). There are about a dozen countable noncontextualized entities - that is, entities which cannot be assumed to be known to the listener - which are introduced by "one", "one lady", "one piece of wood", and less frequently by "a". About fifteen NPs show a somewhat more complex internal organization, for example "the other side", "the father girl"; some of them seem to include a relative clause; we will consider them later.

There are three deictic pronouns ("I", "me", and "you", only in quoted speech, (cf. Ru16 below) and two anaphoric pronouns, "he" and "they"; they refer to Chaplin or to the "father girl." There are no other anaphoric expressions, such as "this" and "that". Finally, Rudolfo uses a dozen prepositions/particles and five connectors: "and", "but", "after", "then", "when".

Lack of space prevents us from giving an utterance-by-utterance analysis of Rudolfo's retelling. To give some idea of his speech, however, there follow two short extracts: part of the shipyard episode, and part of the episode where the girl steals a loaf of bread and is finally arrested. In this transcription, parentheses indicate an unclear word, and the slash / a replanning on Rudolfo's part.

- Ru1. \*e\* the director prison give for Chaplin one piece of paper ... good for job
- Ru2. he go to the factory make the ship
- Ru3. (and) the manager: "its ok"
- Ru4. Chaplin ... take off the jacket
- Ru5. go to work
- Ru6. the boss tell: "give me one piece of wood ..."
- Ru7. Chaplin look for this one piece under the ship
- Ru8. (when) take off the piece
- Ru9. the ship go /
- Ru10. the ship is not finish
- Ru11. go to the sea
  
- Ru12. and the girl ...(is) running ... away

- Ru13. <crash> with Charlie Chaplin  
 Ru14. go to the street  
 Ru15. the policeman take the girl  
 Ru16. Chaplin tell: "is not the girl is me"  
 Ru17. the girl go  
 Ru18. (and) the policeman take ... Chaplin

Ramon, the third informant, comes from Chile, and is in his mid-twenties. After secondary education in Chile, he served an apprenticeship as a joiner. At the time of the interview, he had been in France about fifteen months, the same as Rudolfo in England, and somewhat less than Vito in Germany. His command of the TL is, however, intuitively much better than that of the other informants. This is probably due to two reasons: (1) as a political refugee, he was given a six-month French course on arrival; (2) the special SL-TL configuration "Spanish-French" allows the informant to perceive TL as closely related to SL on the level of lexis and, to a certain extent, of syntax (Noyau, 1984; Giacobbe & Cammarota, 1986). All initial learners in the Spanish-French part of the ESF project develop a vocabulary relatively fast in relation to other learners.

Metalinguistic asides are quite rare in Ramon's case. Quoted speech is frequent, however. It is almost invariably introduced by *dit*,<sup>5</sup> or *demande* (à NP) *qué*. Quoted speech and narration proper do not differ, except for the use of deictic pronouns and the occurrence of questions and commands. So we will include both in the following analysis, and metalinguistic asides will not be considered.

Ramon omitted the shipyard episode in his retelling: We give extracts of the bread-stealing episode, and the episode where Chaplin goes to the restaurant and gets arrested for not paying:

- Ra1. *depuis sé ... une femme qu'el est faim et volé un pain*  
 since (=afterwards) it's a woman, who-she is hunger and steals a bread  
 Ra2. *il sé trouve avec Chaplin*  
 he (=she) is/finds herself with Chaplin  
 (description of Chaplin thinking how nice prison would be)  
 Ra3. *quand sé trouve la femme + pan!*  
 when turns up the woman + pow!  
 (= they collide and fall down)  
 Ra4. *Chaplin il a le pain quand arrive la police*

It is often problematic to assign unambiguous correspondences between Ramon's pronunciation of verb forms and the written French form; for example, /truve/ may correspond to *trouvé*, *trouvait*, *trouver* or *trouvez*. For ease of reading, we succumb to the closeness fallacy ourselves, but note that the verb forms in what follows are highly overinterpreted.

- Chaplin he has the bread when arrives the police
- Ra5. *\*bueno\* la personne qué court devant la fille*  
ok, the person who runs before (=behind) the girl  
*aussi sé trouve avec la police*  
also finds herself with the police(man)
- Ra6. *et lé demande qu'el a volé le pain*  
and to-him asks who-she has stolen the bread
- Ra7. *Chaplin entre un magasin un magasin \*trattora\**  
Chaplin enters a shop a shop \*trattora\*
- Ra8. *parce qu'il mange beaucoup de choses*  
because he eats a lot of things
- Ra9. *\*y\* bon après+il vu passer un police*  
and well after he seen go by a police(man)
- Ra10. *il appelle à le police*  
he calls to the police(man)
- Ra11. *bon et il né lé payé pas à la personne*  
well and he not to-her pays not to the person  
(= he doesn't pay the cashier)
- Ra12. *bon \*y\* la police l'arrête*  
well and the police(man) him arrests
- Ra13. *après dans le voiture il sé trouve avec*  
after in the van he finds/found himself with  
*la fille qué a volé le pain*  
the girl who has stolen the bread

As has been mentioned above, Ramon's vocabulary is comparatively rich, and he is hardly ever in need of a noun, a verb or an adjective. He also has a fairly rich verb morphology, although he is still far from having acquired the TL system. The base form (e.g. /truv/ "find," /envit/ "invite") is still dominant; but there are several infinitives, often with a preposition (e.g. à vivre "to live"), seven clear passé composé, (e.g., Ra13), one clear future, one possible imperfect, and a conditional.

There are two nominative pronouns, *il* and *el*, where both may correspond to standard French *il* or *elle* (Ra1, Ra2); we have transcribed the latter as *el*. This *el* and *les deux* are also used as anaphoric plurals. There is one oblique clitic pronoun, /le/ - transcribed *lé* - which corresponds to standard French accusative (*le*) and dative (*lui*) (cf. Ra6, Ra11). *Lui* "him", *moi* "me", and *elle* "her" are appropriately used as strong forms of these pronouns. There are no other anaphoric devices for NP, except one instance of *ça* "this" and a trace of *ce* in the set phrase /se/ (cf. Ra1) which probably comes from *c'est* "this is."

There are a number of prepositions, *à*, *de*, *dans*, *pour*, *avec*, *devant*, *en face de*, some of them strongly overgeneralized, and several connectors. Most important

among them is /ke/ - transcribed *qué* - which functions as a relative pronoun - both "subject" and "object" - and as a complementizer after verbs of saying (cf. Ra 6). In the former function, the subject relative clause often, but not always, has a resumptive pronoun; compare Ra13 above with

Ra14. *avec le police qué il a tombé ...*  
with the police(man) who he has fallen ...

Ramon makes extensive use of other connectors, as well, such as *et* "and", *mais* "but", *quand* "when", *après* "after", *avant* "before", *parce[ke]* "because", and others; their use sometimes deviates from TL use; thus, his *depuis* (standard French "since") means something like "after" (cf. Spanish *después*, as in Ra1).

### *Influences on Phrasal Construction*

Earlier we listed various kinds of influences or constraints which the learner could follow in order to put his words together. One type was based on simple phrasal conditions, such as "Put the verb at the end" or "Put an NP into initial position." In this section, we will consider influences of this type. It will become clear that there are indeed restrictions statable in these terms but that they do not suffice to account for the learners' utterance structure. They must be completed by other constraints to which we will turn shortly.

Any description of possible phrasal constraints depends on which phrasal categories we assume to exist in the given learner variety. This is no trivial problem, and we will not go beyond the most elementary assumptions. As the detailed analysis of Vito has shown, and as is confirmed by an inspection of the complete data sets, there are at least the lexical categories N, V, Cop(ula), Art(icle), Adv(erb), Pro(noun), Pre(position) as well as the syntactic categories NP and P(rep + N)R Other categories are disputable. Thus, Ramon clearly has complex verbs, consisting of Aux + V, and relative clauses; this is less clear for Rudolfo and unlikely for Vito. In what follows, we will start with the clear categories, and discuss additional possibilities, as they arise.

*The Basic Patterns.* Ramon's constructions have either verbs or copulas and either one or two NPs. This gives us the following six basic patterns:

A1	NP1 V
A2	VNP2
♀	NP1 V NP2
B2	NP1 NP3 V
C1	NP1 (Cop) {PP, Adv, NP2}
C2	{PP, Adv} (Cop) NP2

All constructions may be preceded by a conjunction or some other sentence connector ("then," "now," etc.); all four V-constructions may be completed by an adverbial, that is, a spatial or temporal, sometimes modal, Adv (including,

for Rudolfo and Ramon, "when-clauses") or PR This adverbial is normally utterance-final; it may also appear in initial position, however.

Among all six patterns, A1, B1, and C1 are frequent; A2, B2, and C2 are rare and Rudolfo does not have them at all. Before considering them in more detail, we will first see how NP is expanded. This is obviously different for the three informants. It also depends on where NP appears in the pattern:

<i>Vito</i>	<i>Rudolfo</i>	<i>Ramon</i>
Ø	Ø	Ø
<i>sie</i>	"he", "they"	<i>il, el</i>
<i>diese</i> (N)	-	-
<i>de</i> N	"the" N	<i>le, la, N</i>
<i>ein</i> N	"one/a" N	<i>un, une</i> N
N	N	N
name	name	name
all but	all but	all but
Ø, <i>sie</i>	Ø, "he", "they"	Ø, <i>il</i> plus: Prep <i>lui, el, N</i>
NP3	-	<i>lé</i>

Note that, in Ramon's case, the Prep NP constructions whose NP is a clear argument of the verb (e.g., Ra10: *il appelle à le police*, cf. also the *lé* of Ra11) are assimilated to NP for the purpose of this analysis, limiting PP to cases such as *devant la fille, après l'accident* and so forth. In other words:

1. All informants have three types of lexical NPs<sup>6</sup> that can occur as NP1 and NP2.
2. All informants may have names in NP1 and NP2.
3. All informants have two anaphoric NPs, namely zero and "he" (and equivalents); they occur only as NP1.
4. Vito has in addition *diese* (N), which is anaphorical and may, but need not, have a lexical noun; it occurs as NP1 and NP2.
5. Ramon has anaphoric elements (*lui, el, lé*) as NP2 and NP3, too.
6. Ramon, finally, has a construction not mentioned so far: He may combine NP and *il/el* (e.g., Ra4: *Chaplin il a le pain*) in NP1 position.

Both the similarities and the differences raise interesting questions, to which we will return. Let us consider now the six basic patterns in more detail.

*Verbs with One Argument.* Rudolfo always uses pattern A1 to put an individual NP argument in initial position, whereas Vito and Ramon may also use A2, which puts the verb first. There is a clear condition for use of A2: They are "presentationals," in the sense already discussed, that is, they mark an "appearance on the scene," mostly with the equivalent of the verb "to come"

Lexical NPs may be expanded by some modifier in Rudolfo's and Ramon's case, for example an adjective or even a relative clause. There are also compound NPs, such as "father girl." Since we are not interested in NP-structure as such, and since these cases are rare anyhow, we will not consider them here.

(some of the examples quoted in this section have already been given; for ease of reference, they are repeated here):

V24. *sofort* *komme* *chef* *bäckerei*  
immediately come boss bakery

Ra15. *après* *arrive* *\*otra\** *personne*  
later come other person

This is a first clear case which shows that purely syntactic criteria do not suffice to account for the regularities of utterance structure in learner varieties.<sup>7</sup>

It is worthwhile mentioning that the NP in pattern A1 may play different roles with regard to V. There are at least three types:

1. genuine intransitive constructions, with verbs such as "to go" and their equivalent;
2. "absolute" use of transitive verbs, such as "to pay," where only the agent is mentioned but not what is payed; and
3. "ergative" constructions, such as examples V4 or V21; in this case, no agent is mentioned but a kind of "object," affected by the action, is:

V25. *kleine* *schiff* *mache* (=V4)  
small ship make

Rudolfo has no clear cases of ergatives, Ramon has one where "Charlie" is clearly the topic:

Ra16. *Charlie* + *lé* *doné* *la* *liberté*  
Charlie to-him give the freedom

since this is the first utterance of his retelling and is preceded by a metalinguistic passage where Ramon explicitly asks how one says *\*Carlitos\** in French. Charlie is here the "beneficiary," and the "agent" is left unexpressed, presumably because it is felt by Ramon to be irrelevant.

We mention this heterogeneity of functions in order to stress again that the trivial but tempting rule, "Subject first," simply does not work; there is no or little morphological marking, positional criteria would beg the question (and fail for V NP), and semantic criteria give wrong results in the cases of the quasi-passive and of presentationals.

*Verbs with Two Arguments.* The clearly dominant pattern here is B1: NP1 V NP2. Only Ramon has NP1 NP3, where NP3 is the clitic pronoun *lé*.

Ra17. *la* *police* *l'arrête* (=Ra12)  
the police him arrests

Ramon also has V-NP in subordinate constructions: with *quand* in Ra3 and Ra9. However, it is difficult to tell whether the V-NP order is coincidentally only found in subordinate clauses containing presentational verbs, or whether it reflects the "stylistic inversion" (Kayne, 1972) of spoken French. Perhaps the class of "presentational" verbs will have to be modified in subsequent analysis: Véronique (1985), in a study of a Moroccan acquiring French, notes the V-NP order after verbs such as *marcher* "walk", *monter* "climb", and *partir* "leave" as well.

The distinction between "clitic" and "nonclitic" pronouns offers a straightforward explanation for the different orders NP NP V and NP V NP: It is just like in standard French. But note that this would be a strong (and perhaps fallacious) assumption, since it is arguable that "clitic" in Ramon's variety does not mean the same as "clitic" in the target variety - *le* combines functions that are formally differentiated in French. He uses the corresponding nonclitic forms *lui* and *moi* only with prepositions, and it may well be that this is his criterion. Lack of space prevents us from further pursuing this point. This exception apart, note that B1 still does not solve the whole problem of arranging the words, since it leaves open which NP goes where. As we shall see, the solution must be based on semantic or "pragmatic" criteria.

*Copula-like Constructions.* All informants have a copula. But since they often do not use it where both source and target language would require it (Vito uses it only in quoted speech), we prefer to speak of "copula-like" constructions. They have at least one NP argument, which Rudolfo always has in first position. Vito and - in at least one case - Ramon may have it in final position, too, and the difference is, again, whether it is a presentational or not. The less frequent structure C2 is then the Stative counterpart to A2 (the "arrival on the scene"), and, as with A2, purely syntactic criteria are not sufficient to account for it. In the nonpresentational case, there may be a second ("predicative") NP, and again there is no syntactic criterion to decide which one comes first: With reference to the introductory quotation to this paper, the sequence "\*A linguist is Peter" would be perfectly acceptable in German, for example.

*Summary.* It seems then that the copula, where explicit, behaves like a verb with regard to positional restrictions, and we may sum up all positional restrictions in one rule, where V means V or Cop.

*Rule A: The basic pattern is NP1 V (NP2) except*  
 (a) *in presentationals, which have V'NP2 (Vito, Ramon)*  
 (b) *when one of two NP arguments is an oblique clitic; then the order is*  
*NP1 NP3 V (where NP3 is the clitic NP for Ramon)*

This rule leaves open which NP - if there are two - goes where; it also does not explain the constraints which hold for the occurrence of anaphoric NPs. Finally, with the exception of place adverbials in copula-like constructions, it says nothing about the position of adverbials of place, time, and so forth. As has been said, they are mostly final, but sometimes initial, and we will not try to determine the conditions under which they appear.

*Exceptions.* In a number of cases, it is simply impossible to confirm or disconfirm Rule A; the reasons have been extensively discussed and we will

not take them up again. But there are three clear exceptions, too. First, there are some sentences with more than two overt NP forms: For example:

Ra19. *il ne lé payé pas à la personne* (=Ra11)  
 he (not) it pay not to the person

*Lé* most probably refers forward to *la personne*; thus there are two forms referring to the same argument of *payer*. These cases are too rare, however, to draw any general conclusions.

Second, Vito has at least two NP NP V constructions of the following type:

V26. *Charlie mit de polizei gehe in gefängnis*  
 Charlie with the police go to prison

Actually, the second NP is a PP, but it seems to function here like the second part of a conjunction "Charlie and the police go together ..." - a construction not dissimilar to pidgins and creoles. Again, there are not enough examples to draw any conclusions.

The third exception is more interesting: there are several relative clauses (Ramon) or constructions resembling relative clauses (Rudolfo). Recall that Ramon has a kind of relative pronoun *qué*, which seems to correspond to standard French *qui*:

Ra20. ... *la fille qué a volé le pain* (=Ra13)  
 ... the girl who has stolen the bread

If we analyze *qué* as a regular NP, then Rule A is saved: It also applies to relative clauses. There is one problem, however: Ramon's relative clauses often have an additional resumptive pronoun.

Ra21. ... *avec le police qué il a tombé* (=Ra14)  
 ... with the police who he has fallen

This is a clear violation of Rule A, unless either we analyze *qué + il* on the same level as constructions such as *Charlie il a le pain* ("Charlie he has the bread") or we consider the *qué* in Ra21 to be the trace of a former stage where *qué* is a generalized marker of subordination, thus Ra21 is on a par with:

Ra2. *il pense qué c'est mieux.*  
 he thinks that its better

Rudolfo has no relative pronoun, but some of his constructions resemble a relative clause:

Ru19. One lady tell ... the man work in the backer

Ru20. (Charlie) go to the factory make the ship (=Ru2)

One could save Rule A for this case by arguing that there is a zero-NP functioning as a relative pronoun after "the man" and "the factory", respectively (analogous to standard English constructions such as "there was a man came and asked for Fred"). This is a straightforward and therefore quite tempting description. But, as we have noted, one should be careful with postulating zero

elements unless there is substantial evidence for a clear rule which controls the occurrence of the zero element - and in this case, we have only a couple of examples so far.

### *Semantic Constraints*

If the learner wants to express the fact that Charlie has seen the policeman, Rule A provides him with some, but not with sufficient, information on how to put his words together: It tells him to put his word for "see" between his expression for "Charlie" and his expression for "the policeman"; but it does not tell him which NP comes first. Indeed, the English order expressing the idea "Charlie see policeman" can be expressed in French as "Charlie see policeman," "policeman Charlie see," "Charlie policeman see," "policeman see Charlie," "see policeman Charlie," or "see Charlie policeman." German allows both "Charlie see policeman" and "policeman see Charlie" for the same idea. Therefore, a learner cannot simply associate a fixed position with each verb. There must be additional criteria. In this section, we will consider possible semantic factors.

They may have to do either with inherent semantic properties of the referent - for example, whether the referent is animate, human, or whatever - or with properties relating to the verb or the whole activity, such as agentivity; we will call them "role properties." The example above suggests that inherent properties may be of little help: It is hard to image any semantic feature of either Charlie or the policeman which could serve as a base for their different position. In fact, an inspection of all examples shows that inherent semantic features play at most an indirect role: Animate human NPs strongly tend to occur in first position. But this is simply due to the fact that referents which function as an agent tend to be animate (cf. Silverstein, 1976b). An agentive verb such as "to make" may have nonanimate agents, however, as is illustrated by the following example (=V3):

V27. *diese mache schiff*  
       this (=shipyard) make ship

The crucial semantic factors (if any) are role properties rather than inherent properties. This is clearly corroborated by example V26. If there is a clear asymmetry between the two NPs, for example the one being an agent, the other not, then the former comes first. (We agree that this is not much of a surprise.) The problem is that not all utterances express actions with a clear agent. A principle such as "agent first" does not work for examples like the following ones (all taken from Vito):

V28. *Ich brauche eine keil* (=V7)  
       I need a wedge

V29. *Ich wolle eine keil* (=V12)  
       I want a wedge

V30. *Sie* *habe* *brief* (=V1)

She (=Charlie) has letter

V31. *Diese* *holz* *sicher* *schiff* (=V20)

This wood safe (holds) ship

One might certainly argue that the relation between the two NPs in V31, though static, is more actionlike than, for example, the one in V29, and the "log," although not an agent, is at least more agentive in V31, when compared to the other NP referent "ship," as we have already argued above. So, one might replace categorical distinctions such as "agent" and other "case roles" by a scale which also extends over nonagentive relations. Looking through our examples, it looks as if the "degree of control" might provide us with such a scale: It reflects the degree to which one referent is in control of, or intends to be in control of, the other referents. The degree of control varies with the (nonnegated) relation: thus, "to make" provides us with a stronger control asymmetry than, for example, ownership, as in V30, or intended ownership, as in V28 or V29. But in all of these cases, it gives us a semantic role asymmetry - an NP referent with "higher" intended control and another with "lower" intended control (for related insights, see the well-known studies of Hopper & Thompson, 1980; Silverstein, 1976b). This allows us to state the following rule:

*Rule B: The NP referent with highest control comes first.*

Admittedly, the relational property "being in control of" needs a more precise characterization, for example in terms of verb classes. But this being granted, Rule B solves a great deal of arrangement problems in sentences with two NPs. It does not work, however, for verbs which do not convey a control relationship such as the copula with two NPs, or uses of the verb "have" for property assignments (the sequence "\*one handle has this cup" is very natural in German). The relation "x is y" seems to provide no semantic asymmetry. Hence, the difference between "The girl is the thief" and "The thief is the girl!" - must be due to other factors, to which we will now turn.

### *Pragmatic Factors*

There are two arrangement problems left, which cannot be accounted for by Rules A and B. These are the symmetric copula constructions and the specific occurrence restrictions of NP types: Anaphoric NPs, including zero, are subject to specific positional constraints. We will start with the latter problem.

As we have seen in the discussion of Vito, and as one would expect to find, the occurrence restrictions of these various NP types are closely related to whether a referent is first introduced, reintroduced, or maintained from some preceding utterance. An inspection of all examples leads to the following - quite straightforward - conclusions:

1. "the" + N (and equivalents) and names are used when the referent can be assumed

- to be known to the listener, either because it was referred to before, or because it is associated with some entity referred to before, or because it can be assumed to be part of the listener's general knowledge;
2. "a" + N (and equivalents) is used for first introduction of a referent.
  3. The use of bare N is not totally clear. It is often used to introduce or to maintain noncountables, but there are frequent exceptions. It may well be that the use of bare N reflects previous learning stages (for a more detailed investigation of this problem cf. Carroll & Dietrich 1985).
  4. Zero and "he" are used to maintain a previously introduced referent (so do *diese* + N in Vito's case, and *lé/lui* in Ramon's).

This leaves us essentially with one important question: What causes the different types of maintenance - name, "the" + N, zero, "he" and equivalents? In Ramon's case, there is also the question of the conditions under which he uses NP + *il*.

In all cases, the referent has already been introduced. The difference between them apparently has to do with the nature of the referent (in these texts, human or not), and the position where it was previously referred to: in which utterance, and where in this utterance. We illustrate this with the first three utterances of Vito's text V1, V2, and V3 which we repeat here.

V32. (Charlie is introduced in the previous discourse, which is not part of Vito's narration).

*sie habe brief..*

*0 komme in eine baustell*

*diese mache schiff*

For zero to apply, two conditions must be met. Condition A is that the referent must be maintained from the immediately preceding utterance of the narrative text, and Condition B is that the referent must be in initial position. If Condition A is not met, *sie* must be used; if Condition B is not met, Vito uses *diese* (we will come to the other informants shortly). In practice, however, Condition A is often violated in that zero may jump over intervening utterances; but then, these utterances do not belong to the "plot line" of the story; they may give background information or be metalinguistic comments. In order to make Condition A more precise, we have to account for this difference in pragmatic function within the text. Also, Condition B seems to reflect some functional difference: Intuitively speaking, zero and (perhaps) "he" seem to require "topic maintenance," whereas the switch from *in eine baustell* to *diese* seems to reflect a transition from "focus" to "topic." Now, as we stated earlier, all of these terms are highly disputable. In what follows, we will work out a simple proposal to account for these problems in the present context.

Very often, a statement is used to answer a specific question, this question raising an alternative, and the answer specifying one of the "candidates" of that alternative. For example, the question "Who won?" raises an alternative of "candidate" persons - those who may have won on that occasion, and the

answer specifies one of them. A question may raise all sorts of alternatives, for example, actions ("What did Charlie do?"), contents of prepositions ("Was Charlie *before* or *behind* the ship?"), etc. Let us call "focus" that part of a statement which specifies the appropriate candidate of an alternative raised by the question, and "topic" the remainder of the answer.

Now, not all texts are question-answer-sequences. But we may assume that any statement is an answer to an (implicit or explicit) question, which we will term *questio* to remind the reader that it is an analyst's construct. Thus, Vito's *Ø komme in eine baustell* is an answer to a (implicit) *quaestio* "What did Charlie do at that time?" whereas *diese mache schiff* answers an (implicit) *quaestio* "What did this baustell do?" Note that the two statements serve quite different functions within the whole narrative: The first indeed belongs to the "plot line" - the foreground of the story - whereas the second gives (relevant) background information: it does not answer the "key question" of the whole text, which is: "What happens with Charlie (and possible other protagonists)?" Thus, all utterances which answer the key *quaestio* belong to the foreground - they push forward the plot line - and all other utterances, no matter which (possibly very important) *quaestio* they answer belong to the background.

A narrative is an answer to a question-function  $Q_i$  where  $i$  ranges over time intervals:

- Q1: What happens with  $p$  at  $t_1$   
 Q2: What happens with  $p$  at  $t_2$   
 Q $_n$ : What happens with  $p$  at  $t_n$

This question-function defines the foreground of the narrative:<sup>8</sup> all utterances which are answers to one of these questions are "foreground-utterances". It also defines topic and focus within each of these utterances: The topic of a foreground clause includes a time span  $t_i$  (which is mostly not explicitly specified but given by the sequential order, except for  $t_1$ ) and a protagonist or the protagonist. The focus specifies the action or event at that time span (which means, incidentally, that foreground clauses normally cannot be imperfective or stative: One could characterize exceptions to this "norm" to a certain extent, but lack of space prevents us from doing so). The focus specifies then a possible "happening" at that time  $t_i$  with that protagonist  $p$ . Background clauses are normally linked to a foreground-clause; their internal focus-structure is quite different, depending

Narratives may differ to some extent with respect to the "key question". We have chosen the relatively neutral formulation "happens with  $p$ ," although it gives the protagonist (or protagonists) a somewhat passive role. Let us add that it might be more appropriate to characterize a narrative by two "key questions" ( $Q_0$ ,  $P_1$ ), where  $Q_0$  refers to the "rooting" of the whole event in time and space (Labov's orientation): "When and where did  $a$  happen?" where  $a$  is the total event (one may indeed imagine a third "key question": "so what?"). This whole approach, which also applies to other types of texts, is worked out in more detail in Klein and von Stutterheim (1987); here we give only the rough idea.

on what information they specify. The overall structure of a narrative is then (A = foreground, B = background):

Q1          Q2          Q3          Q<sub>n</sub>  
(B) - A<sub>1</sub> - (B) - A<sub>2</sub> - (B) - A<sub>3</sub> - (B) - A<sub>n</sub> - (B)

This structure plays an important role in the narratives of all informants. We noted that Rudolfo has a base form and a base + "-ing" form for V. It turns out that the latter massively occurs in B-clauses. Similarly, all "subordinate clauses" in Rudolfo's and Ramon's texts are of type B. There are also some immediate implications for word order to which we will turn in a moment.

Using the abbreviations T for topic and F for focus, we may now restate our observations in connection with V1-3 (repeated in V32 above): Zero maintains a referent from T in A<sub>i</sub> to T in A<sub>i+1</sub>, *sie* a referent from somewhere to T in A<sub>i</sub>, and *diese* (N) from F in A<sub>i</sub> to T in A<sub>i+1</sub>. In other words: What matters for the use of the various types of NP is not only whether something is maintained or introduced, but also whether it goes from T to F (topic to focus), from F to T and so forth, and which clauses intervene.

An analysis of all of the texts shows the following regularities:

1. Transition from "nothing" to T (=first introduction): lexical NP (except *diese* + N in Vito's case). Note, however, that "the" is quite rare.
2. Transition from T to T ("topic maintenance"): zero, "he," "the" + N; zero only applies when the two clauses are adjacent, and where B-clauses do not normally interrupt adjacency of A-clauses. The difference between "he" and "the" + N is not totally clear; it seems, however, that "the" + N is used when there might be an ambiguity, e.g. when there is an intervening NP which could be misinterpreted as coreferential to "he."
3. Transition from F to T: "he" and, in Vito's case, *diese* + N.
4. Transition from "nothing" to F ("focus introduction"): lexical NP (or name), except *diese* (N); again, "the" + N is rare (it is only used when the referent can be assumed to be known, although it was not mentioned before).
5. Transition from T to F: There are few clear examples, probably lexical NP (or name) for Vito and Rudolfo as the following two utterances show:

Ru24. one man ... go to the work  
(his) wife kiss the man.

6. Transition from F to F: few clear examples, probably lexical NP (or name). (Note that this case *is* possible, as can be seen with *un police - le police* in Ra9 and Ra10.)

This leaves us with Ramon's *Charlie + il* construction. This pattern consistently serves to reintroduce a topic: It "highlights" that there is a new topic with respect to the preceding clause; but that the referent of this topic has already been introduced into the discourse.

So far, we have used the T/F distinction only to describe the regularities in the use of various types of NR. But apparently, it also has a direct bearing on the word arrangement problems. The crucial rule is quite simple:

*Rule C: Focus comes last.*

This rule is gross. In particular, it does not take into account the fact that both the topic-component and the focus-component usually have an internal organization. Still, it helps to answer some open questions. One of them is the "copula-problem." If a copula has two NP arguments, then there is no semantic criterion to decide about their order, as in the case of NP V NP. Here, Rule C puts that NP which is (or belongs to) F at the end. Rule C also gives a possible explanation of the word order in Ramon's and Vito's static (Cop NP) and dynamic (V NP) presentationals: It seems plausible to assume that in presentationals, the NP is the answer to the (usually implicit) *quaestio*, and hence, it comes last. Rudolfo, on the other hand, sticks to the strictly phrasal principle, "NP before V," which, defined in purely categorial terms, wins out in the competition with the pragmatic Rule C. And finally, Rule C provides us with an explanation of a problem which was mentioned only in passing: the position of (temporal and spatial) adverbials, which may appear in initial or in final position; as an inspection of the examples shows, this difference in position is clearly related to their function as part of the topic-component ("orientation") or as part of the focus component.

Summing up, we have found that three "pragmatic" factors play a role for the utterance organization in learner varieties:

1. Familiarity: Can a referent be assumed to be known to the listener, either by world knowledge, or by contextual information of various sorts?
2. Maintenance (vs. introduction): Is a referent first mentioned, or was it already referred to in a previous utterance?
3. Topic-focus structure: Does a constituent specify a candidate of the alternative raised by the *quaestio* of the utterance?

## Conclusions

The objective of this paper was to analyze whether there are any principles according to which learners with a limited repertoire put their words together. It was shown - with some exceptions and some degree of uncertainty - that there are basically three rules which determine the arrangement of words in early learner varieties (plus one rule for the type of NP which may occur in a specific position): a phrasal, a semantic, and a pragmatic rule. In other words: All of the three possible kinds of constraints suggested in I, II, and HI of the introduction indeed play a role. What we have not considered so far, is their possible interplay. What happens, for example, when Rule B, "low control unit last," and Rule C, "F last," are at variance, when, for example, a clear agent is in F? Apparently, our informants were quite skillful in avoiding such conflicts: There are few clear examples which would show how one factor is outweighed by another one.

But generally speaking, whereas Ramon and Vito allow for more pragmatic and semantic control of word order, as (spoken) French and German in general

do when compared to English, Rudolfo favors a "coalition" of phrasal and semantic constraints: The NP is preverbal in presentationals, and the agent is preverbal in utterances with action verbs (the majority), which results in more frequent transitions from T to F, and F to T, than with the other two informants.

This brings us to a last point. We discussed some possible sources of word arrangement principles: source language, target language, and various kinds of universals. It seems clear by now that none of the three rules comes from a syntactic universal; Rules B and C possibly reflect cognitive universals. Where does Rule A then come from? There is no convincing answer at this point; but it corresponds neither to the phrasal constraints of Italian or Spanish, nor to those of the target languages except perhaps for English. It appears to be a genuine syntactic constraint on the three learner varieties considered here, resulting from the learners' interpretation of the input on the basis of their available linguistic knowledge.