

The psychological reality of syntactic dependency relations

Richard Hudson

Department of Phonetics and Linguistics - University College London

Gower Street, London, WC1E 6BT

dick@ling.ucl.ac.uk

Mots-clefs – Keywords

dépendance, psycholinguistique, traitements, prototypes, apprentissage, esprit
dependency, psycholinguistics, processing, prototype, learning, mind

Abstract - Résumé

I review some evidence that dependency structure is psychologically real (and by implication that phrase structure is not), I start by comparing WG dependencies with MTT surface syntax and deep morphology. I then consider the following kinds of evidence:

- Dependency distance: dependencies tend to be short, and longer dependencies tend to be harder to process.
- Dependency direction: consistent direction (head-initial or -final) is cognitively simpler but harder to process, so mixed directions are functionally motivated.
- Dependency classification: grammatical functions are sub-types of dependency, and require the same kinds of cognitive machinery as non-linguistic relations such as kinship relations.
- Dependency prototypes: dependency types (e.g. 'subject') are prototypes which combine observable and unobservable features in bundles which allow exceptions.
- Dependency parsing: dependencies are added one at a time to the head word, rather than by the addition of extra 'mother' nodes.
- Dependency lexicalisation: many dependencies are stored as relations between lexical items.
- Dependency learning: every dependency can be learned by induction from adjacent word pairs, and its properties can be elaborated through experience.

Je passe en revue certains éléments qui prouvent que la structure de dépendance est psychologiquement réelle (et par implication que la structure syntagmatique ne l'est pas). Je commence par comparer les dépendances dans le formalisme WG avec la syntaxe de surface et la morphologie profonde de la théorie Sens-Texte. Ensuite j'envisage les types de preuves suivants.

- Distance des dépendances : les dépendances ont tendance à être courtes, et les dépendances plus longues ont tendance à être plus difficiles à traiter.
- Sens des dépendances : une direction affirmée (tête-initiale ou tête-finale) est cognitivement plus simple mais plus difficile à traiter, donc des directions mélangées sont fonctionnellement motivées.
- Classification des dépendances : les fonctions grammaticales sont des sous-types de dépendance, et requièrent les mêmes types de machinerie cognitive que des relations non linguistiques comme les relations de parenté.
- Prototypes des dépendances : les types de dépendance (par exemple, "sujet") sont des prototypes qui combinent des caractéristiques observables et non observables dans des paquets qui admettent des exceptions.
- Analyse des dépendances : les dépendances sont ajoutées une par une à la tête, plutôt que par l'addition de noeuds pères supplémentaires.
- Lexicalisation des dépendances : de nombreuses dépendances sont stockées comme des relations entre des éléments lexicaux.
- Apprentissage des dépendances : toutes les dépendances peuvent être apprises par induction à partir de paires de mots adjacents, et leurs propriétés peuvent être élaborées par expérience.

1 Dependency in MTT and in WG

One of the main similarities between the Meaning-Text Theory (MTT) and my theory, Word Grammar (WG)¹, is the assumption that syntactic structure consists of dependencies between individual words. For example, consider sentence (1),

(1) Word Grammar structures consist of dependencies between words.

Most current theories of syntax would give this a 'top-down' analysis in which the sentence is broken down into phrases, which are successively broken down until the smallest units are reached, as shown in phrase-structure trees like Figure 1.

In contrast, dependency analysis shows only the relations between individual words - for example, the relations between *Word* and *Grammar*, between *Grammar* and *structures*, and between *structures* and *consist*. This is the approach found in the various theories of grammar which

¹Word Grammar has a web-site (<http://www.phon.ucl.ac.uk>) which gives a great deal of information about the theory, including a number of downloadable introductory papers. A good place for an experienced linguist to start might be the introduction for postgraduate students. There are also brief introductory articles in the big encyclopedias of linguistics (Hudson, 1994; Hudson, 2003), and two book-length discussions (Hudson, 1984; Hudson, 1990) as well as numerous journal articles.

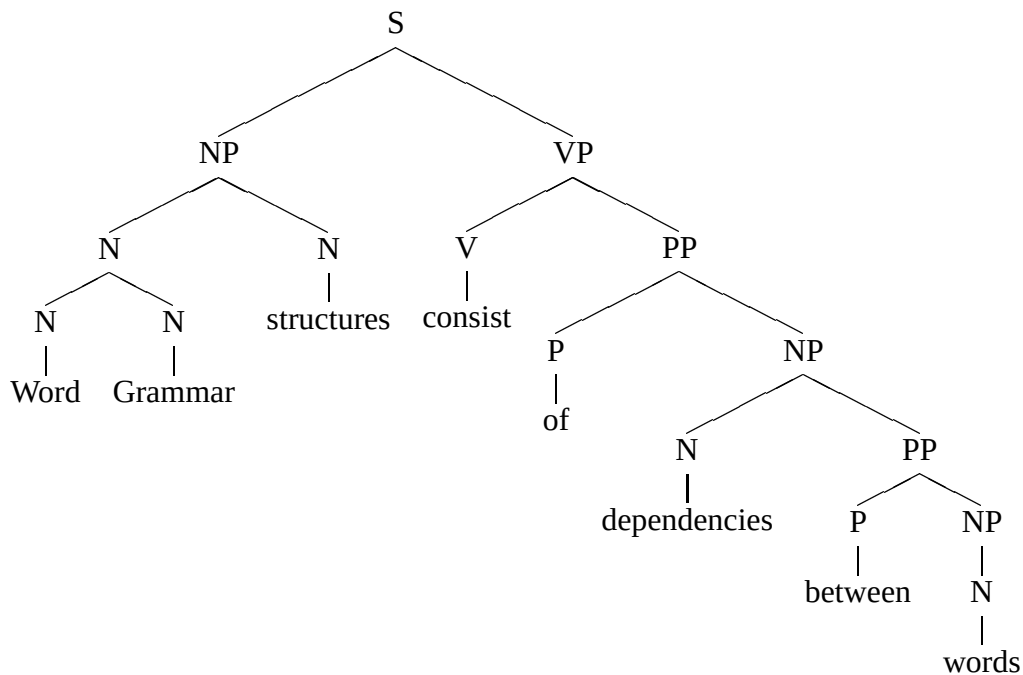


Figure 1: An elementary phrase-structure tree

belong to the general family called 'Dependency Grammar'², whose principle members include MTT and WG. However there is another important point of agreement among dependency grammarians: these relations between pairs of words are asymmetrical. (The obvious exception is coordination, but we all agree that some kind of special arrangement has to be made for this.) One word is always subordinate (in some sense) to the other. Traditionally this is shown by means of the stemma (as proposed by Tesnière), and this notation is commonly used in MTT for syntactic and semantic dependencies. In a stemma, words are connected directly by a line whose slope shows which word depends on the other and whose label shows the exact nature of the relationship. I guess an MTT surface-syntactic analysis would be something like Figure 2

The structure in Figure 2 is order-neutral, since linear order is shown only in the deep morphological structure, but as Kahane shows (Kahane, (forthcoming)) the two can be combined into a single structure which puts the words in left-right order and links them by arcs that carry labels. This is precisely the same notation as I use, so we can see how easy it is to translate between MTT and WG. In short:

WG syntax = MTT surface syntax + deep morphology . (It's actually a bit more complicated than that, but we can leave it simple for the time being.) Figure 3 is like the one that Kahane suggests, but its labels come from WG rather than MTT.

This similarity between WG and MTT is encouraging given that the two theories were developed more or less independently of each other. Of course there are important differences as well - differences in the details of analysis as well as in the overall architecture of the system. But the central role of word-word dependencies in syntax is quite striking.

²For more information on Dependency Grammar, visit <http://ufal.mff.cuni.cz>.

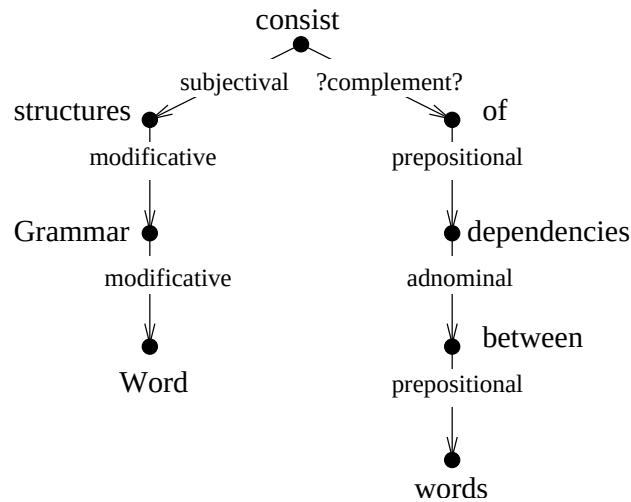


Figure 2: An MTT surface-syntactic analysis

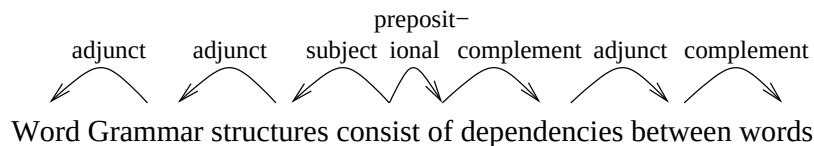


Figure 3: A WG syntactic analysis

2 Evidence for dependency structure

The similarity just noticed is all the more encouraging because the two theories have different goals. As I understand it, MTT is a formal theory of language developed within the context of machine translation and descriptive linguistics, whereas WG is more oriented towards human psychology. WG tries to model human competence in the light of what we know (from psychology) about human performance, which puts it in the recent tradition called cognitive linguistics. The fact that both MTT and WG prefer dependency structure in spite of these different goals shows that one's choice between dependency structure and phrase structure is not driven by one's aims. Maybe it really is a matter of truth. If dependency structure turns out to be more useful in computational linguistics, maybe that's because it's nearer to the true nature of language.

My aim in this paper is to review some evidence that encourages me to believe that dependency structure is not only useful for descriptive and computational work, but is psychologically 'true'. By this I mean that dependency structures like Figure 3 are better models for the way we represent sentences in our minds than the phrase structures like Figure 1. In each case the evidence suggests that:

- we **do** pay attention to the relations between individual words, even when they are embedded in larger phrases;
- we do **not** pay attention to the phrases that these words compose by virtue of their dependency relations.

It is important to be clear about these claims, because those of us who work within the dependency tradition are accustomed to misunderstanding. We do not believe that words are put together without structure, like beads on a string; we do believe that they combine to form phrases; but we also believe that the phrases are a mere by-product of the basic word-word dependencies. A dependency structure implies a phrase structure, but the phrase structure adds nothing to the dependency structure. This claim has been discussed and justified at length in the dependency literature, not least by Igor Melčuk (Melčuk, 1988; Melčuk, 1997), so I shall simply take it for granted. However the debate so far has tended to focus on the data of normal linguistics - facts about languages and arguments about how to describe them in a revealing way. In this paper I shall focus on a range of evidence that, so far as I know, has not been generally recognised as supporting dependency analysis.

3 Dependency distance

One of the most striking facts about Figure 3 is that each word is immediately next to the word on which it depends. If we use the term 'dependency distance' as a measure for the length of each dependency arrow, and if we simply count the number of words between the dependent and its 'parent' (the word on which it depends), we can say that each word in this sentence has a dependency distance of 0. This is an important measure for the psychology of processing because it indicates how long the earlier word must be kept 'active' in working memory, which is a measure of how much load the word places on the hearer's (or speaker's) memory. Obviously, the lower the distance, the lower the processing load.

Let me explain my assumptions. I assume that a word must remain active until the hearer has finished processing it, which means until it has been integrated semantically with other words. In dependency terms, this means until it has contributed its meaning to that of the word on which it depends. For example, consider sentence (2), in which dependency distances are rather great:

(2) If when you read this sentence you can understand it easily you must be a genius.

Figure 4 gives the bare bones of a WG analysis:

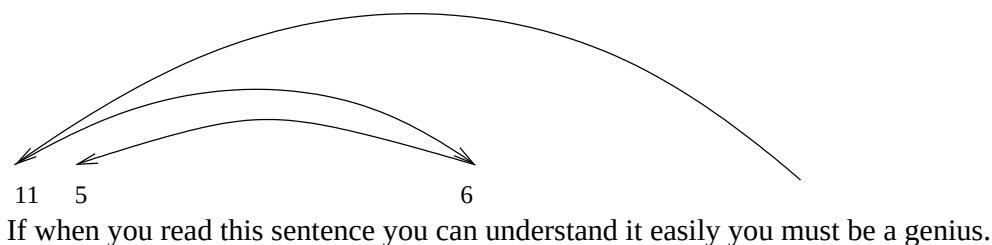


Figure 4: A WG analysis showing some dependency distances

As you can see, the word *if* has a dependency distance of 11, which means that it must be kept active while 11 other words are processed - including two which themselves have relatively large distances.

The notion of dependency distance is thus a simple predictor of processing difficulty, and provides an alternative to measures based on phrase structure such as Frazier's weighted node-

counting (Frazier, 1985) or Gibson's Syntactic Prediction Locality Theory (Gibson, 1998). It is a much simpler and more direct measure precisely because it is based on word-word dependencies and ignores phrase nodes. If sentence structure really was based on phrase structure, there would be no reason to measure the distance between the head-words of related phrases; but in dependency structure the head-words are the only psychologically real elements.

Moreover dependency distance allows two generalisations which strike me as potentially important:

- Languages evolve in such a way as to favour structures which minimise dependency distance - compare similar suggestions by Hawkins (Hawkins, 2001; Hawkins, 1994). The most obvious consequence of this tendency is the observed preference for languages either to be consistently head-final (like Japanese) or consistently head-initial (like Welsh); for example, if a head-initial phrase is part of a head-final one, inevitably the two head-words must be separated by other words.
- Users tend to choose structures in such a way as to minimise dependency distance. More specifically, casual conversation tends to have an average dependency distance of about 0.3. Putting this figure differently, about 70% of words in casual conversation are next to the word on which they depend, and most other words are separated from it by only one other word. This figure has been found independently in numerous undergraduate assignments on English, and also in a careful study of Japanese conversation (Hirayama, 1999; Hirayama, 2001). On the other hand slightly higher figures have been found for German (by Eva Eppler in unpublished work), so there may be some cross-language variation.

If dependency structure really is what we have in our minds, dependency distance is a more natural way to measure processing load than any measure based on phrases. The results from research based on this measure support this conclusion.

4 Dependency direction

In the previous section I referred to "the observed preference for languages either to be consistently head-final (like Japanese) or consistently head-initial (like Welsh)". As we all know, this preference is actually rather limited, and is invisible in all the many languages like English and French which favour SVO word order and in which some dependents stand on one side of the head word and others stand on the other. Fortunately, dependency distance actually explains why languages favour head-medial order (SVO) even more than consistent head-initial order (VSO). This is because of the problem that arises when a word has **two** dependents.

The problem is obvious, and can be seen in Figure 5. This presents the three ways in which a word's two dependents can be ordered: both before it, both after it, or one on each side. In the first two cases one dependent is inevitably separated from the head word, and the separation causes processing problems if the separating dependent itself has many dependents. (One way for a language to react to this problem is by making dependencies optional, as in Japanese; this is why the dependency-distance figure quoted above was so low.) In the third case, on the other hand, both dependents may be next to the head word, and processing is facilitated.

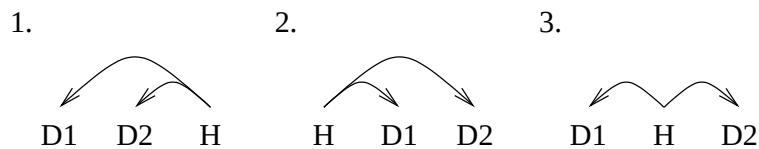


Figure 5: the three ways of arranging a word and two dependents.

This abstract discussion illustrates clearly the processing advantages of a 'mixed' word order, including the SVO order of clause structure. Indeed, if you look at the grammar of English, it is not just the dependents of a verb that are divided as in arrangement 3 - all kinds of words have dependents both before and after:

- nouns (e.g. *big books about ...*)
- prepositions (e.g. *directly under it*)
- adjectives (e.g. *very keen on ...*)

English, then, is a consistent language after all - consistently mixed. It seems that there may in fact be a tendency for languages to be consistent in choosing one of the patterns in Figure 5 for all constructions. If so, this is easy to explain in psychological terms, but the explanation has everything to do with the distance between individual words, and nothing to do with any measure based on phrase structure.

5 Dependency classification

One of the great advantages of dependency structure is that it defines the general notion 'dependent'. Even those who use phrase structure talk informally about heads and dependents (e.g. (Huddleston, 2002):24), but these notions are actually alien to phrase structure. The mainstream view in theory-driven syntax is based on X-bar theory, which recognises three distinct non-head functions - complement, adjunct and specifier - but each is defined by a specific place in structure, and there is no super-category that brings them all together. Thus it is really only in dependency grammar that it is possible to formally express the generalisations about word order that I discussed in the previous section.

But if 'dependent' is indeed a super-function, then all the grammatical functions must be subtypes of 'dependent'. This is again explicit in Huddleston and Pullum's recent work, but it has nothing to do with their phrase structure. In contrast, dependency theories such as MTT and WG build directly on the general/specific relation by showing the super-function as an arrow and its subcategories as labels on the arrow, as in all the examples given so far. Indeed, WG goes further by recognising an extended hierarchy of grammatical functions, whose apex is the super-function 'dependent'. Part of the hierarchy for English is shown in Figure 6.

In this diagram the little triangle is the standard WG notation for the 'isa' relation: valent isa (i.e. 'is a') dependent, complement isa valent, and so on. Some of the terminology may be unfamiliar or confusing if you're used to MTT, and the logic may be puzzling if you're not used to default inheritance (the basic logic of WG), but neither problem should obscure the general

point, which is that dependencies can be sub-classified. Incidentally, the contrast between pre- and post-dependent is the mechanism for handling word order in a consistently mixed-order language.

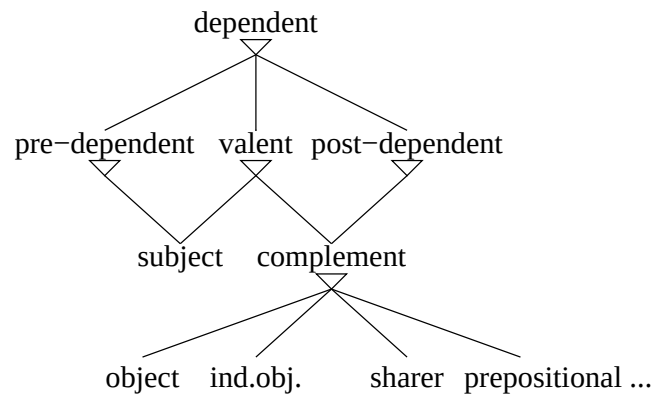


Figure 6: part of the hierarchy of grammatical functions for English

How is this relevant to psychological reality? The point is that this classification of dependents must be part of our mental grammar for exactly the same reason that our mental grammar must contain other classifications - e.g. classification of words as nouns, verbs and so on. As speakers we treat subjects differently from objects, so we must recognise this distinction in our minds. But if we sub-classify dependencies, the super-category 'dependent' must also exist in our minds, because otherwise it could not be sub-classified.

Incidentally, there is nothing psychologically strange or difficult about classifying relations, because we do this very successfully in our social life; just think of your mental system of kinship relations within the family, as reflected by terms such as parent, brother, aunt. Given that we classify (mentally) these social relations, and presumably many other kinds of non-linguistic relations in our experience, it would be surprising if we did not do the same in syntax.

6 Dependency prototypes

Another argument for the psychological reality of dependencies is that they are just like other mental categories. It is widely agreed that mental categories such as 'bird' are defined round clear cases, or 'prototypes' (Rosch, 1976); for example, the prototype bird is (at least in the UK) like a robin or blackbird, rather than a swan. This organisation produces the 'prototype effects' in classification whereby some examples may be either poor examples of the category (e.g. swans and ostriches are bad examples of birds) or borderline examples (e.g. ash-trays are borderline examples of furniture). The point is that mental categories are 'cluster concepts' in which a number of different characteristics typically cluster together, but sometimes one or two characteristics are missing or exceptional and none are clear badges of membership.

The grammatical functions, which I have just suggested are sub-categories of dependents, are very typical cluster concepts (Taylor, 1995). For example, it is generally accepted that grammatical subjects typically share a large number of features universally, and also in particular languages (Keenan, 1976). In English, the typical subject has the following characteristics (among many others):

- It precedes the verb.
- The verb agrees with it.
- It is a noun (or pronoun),
- It defines the 'subject-argument' (another cluster concept) of the verb.

The subject of (3) is typical, but those of (4) and (5) are not:

(3) **John** loves Mary.

(4) Here comes **John**.

(5) **There** is a fly in my soup.

No doubt we could demonstrate the same kind of patterning for all the other dependency categories.

What about the notion 'dependent' itself? This too is a cluster concept. We could all produce a list of characteristics found in a typical dependent; for example:

- It stands next to its parent (the word on which it depends),
- It modifies the meaning of its parent.
- It is selected by its parent.
- It needs its parent.
- It only has one parent.

A typical dependent is a verb's object, such as *Mary* in (3) above, which has all these characteristics. But we can all think of problematic cases where it is unclear whether or not there is a dependency relation. For example, take German noun phrases like (6) and (7),

(6) der gute Mann, 'the good man'

(7) ein guter Mann, 'a good man'

Notice how the inflection of the adjective 'good' changes when the determiner changes from 'the' to 'a', suggesting some kind of selection relation between the determiner and the adjective. This suggests that the adjective depends on the determiner, but if that is correct, the dependency is untypical because the adjective appears to depend on the following noun as well.

What this discussion has shown is that the category 'dependent' and all its sub-categories are typical cluster concepts. But if that is so, they must be concepts - i.e. they must be psychologically real. In contrast, I know of no such evidence for the reality of phrasal categories.

7 Dependency parsing

'Parsing' is the technical term for syntactic analysis either by a machine or by a human. Here we are concerned with how we humans understand the syntax of the sentences we hear. I have

already discussed the question of memory load, which is an important issue in parsing theory. In this section I shall consider how we build dependency structures, and I shall show that this target is much easier to achieve than the target of building phrase structure. The conclusion will be that our minds are more likely to look for dependencies than for phrases.

Suppose you are trying to understand sentence (8),

(8) John loves Mary deeply.

A successful outcome is the structure in Figure 7, where I offer the choice between a typical (but simplified) phrase structure and a dependency structure.

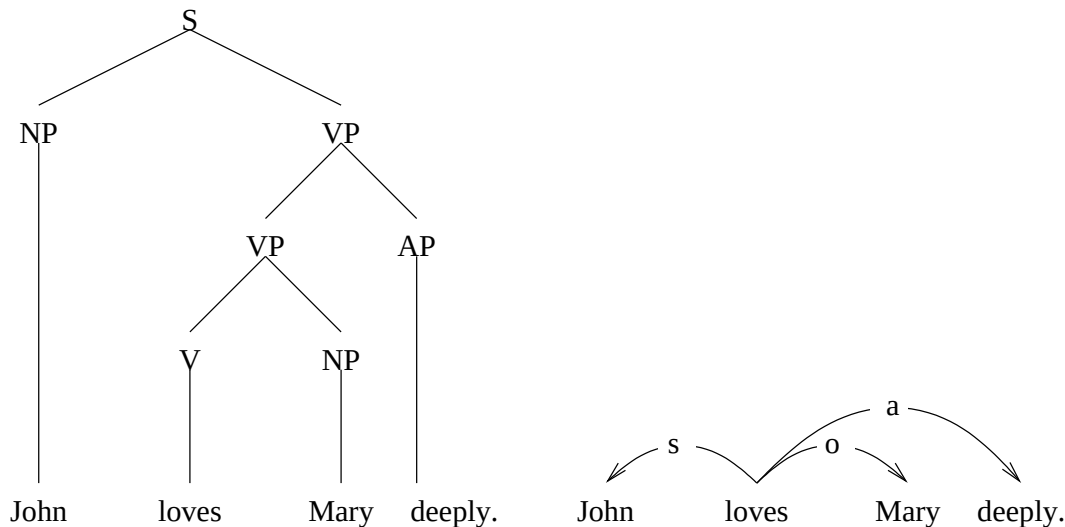


Figure 7: a phrase structure and a dependency structure

It is probably obvious which structure is easier to build - if not, just count the number of nodes which have to be recognised and classified. However there is also a more subtle advantage in aiming at a dependency structure which emerges if you track the structure as it grows through time. At the point where the parser has heard just the first three words - John loves Mary - it does not know that deeply will follow. For the dependency structure this does not matter, because the extra structure for deeply is just an addition to what already exists. But it does matter for the phrase structure because the parser has to change what already exists by adding an extra node - another VP node - to accommodate the new adjunct.

The example is very simple, and very typical. The advantage of dependency structure is that it shows relations between words directly, as relations between words. This is ideal for a parser (human or machine). In contrast, phrase structure shows relations between words indirectly via the phrase structure which holds them together, which means that the parser has to constantly adjust structures to take account of new words. It seems unlikely that our minds would prefer this more complicated option.

8 Dependency lexicalization

This is a very simple point: a lot of words select particular lexical items as their dependents, so these dependents must be single words rather than phrases. Moreover, since we know these

restrictions, our knowledge must be based on word-word dependencies rather than on phrase structure. For example, we know that the verb *depend* takes the preposition *on*, that *cope* takes *with*, and so on (and on),

These facts are easily stored in our minds as facts about individual words - the complement (more precisely, the 'prepositional') of *depend* is *on*, and so on. But they are much harder to express as facts about phrases, because it is not enough to know that *depend* takes a prepositional phrase - it needs a prepositional phrase whose head word is *on*. The facts can be accommodated in phrase-structure analysis only by providing extra apparatus (such as a separate syntactic feature for each preposition).

9 Dependency learning

Dependencies are easy for a child to learn from experience because they can all be found between adjacent words. (At least this seems to be true of English.) Indeed, what I said about dependency distance in ordinary conversation shows that the vast majority of dependency tokens are in fact between adjacent words. We know that children learn syntactic patterns by induction from specific lexical combinations - e.g. a child knows pairs like *eat yogurt* and *kick ball* before it induces the general pattern of verb + object (Tomasello, 2000). If syntax is based on dependencies, this learning is easy to explain: the child remembers recurrent pairs of adjacent words, notices that one modifies the meaning of the other (e.g. eating yogurt is a kind of eating, not a kind of yogurt), and thereby learns that one depends on the other. The minority of cases where the dependent is separated from its parent simply don't count - the child doesn't remember them, and maybe doesn't even understand them. In this model, dependency is simply a generalisation across memorised adjacent word pairs.

In contrast, phrase structure implies a much more complicated process in which the child has to impose phrase structure on the observed words; for example, the only way to show that *eat* and *yogurt* form a unit is by recognising a phrase node that contains both. To return to the comparison with social relations, this is like assuming that the child can only recognise its relations to its mother by recognising itself as a member of a mother-child pair. This seems unnecessary, and unlikely to be true.

10 Conclusion

I have presented various kinds of evidence in favour of the hypothesis that dependency structure is psychologically real and phrase structure is not. If true, this conclusion is important because very few psychologists and psycholinguists pay attention to the dependency tradition in syntax, and some may not even be aware of this tradition. This is partly because dependency grammarians themselves have been less interested in psychological issues than in language description and formal computational applications, but I hope we shall see a change over the next few years.

References

- Frazier L. (1985), Syntactic complexity, In Dowty, D. R., Karttunen, L., & Zwicky, A. M.(eds.), *Natural Language Parsing. Psychological, Computational and Theoretical Perspective*, Cambridge, Cambridge University Press, 129-189.
- Gibson E. (1998), *Linguistic complexity: locality of syntactic dependencies*, Cognition, Vol. 68, 1-76.
- Hawkins J.A. (2001), Why are categories adjacent?, *Journal of Linguistics*, Vol. 37, 1-34.
- Hawkins J.A. (1994), *A performance theory of order and constituency*, Cambridge, Cambridge University Press.
- Hiranuma S. (1999), Syntactic Difficulty in English and Japanese: a Textual Study, *UCL Working Papers in Linguistics*, Vol. 11, 309-322.
- Hiranuma S. (2001), *The syntactic difficulty of Japanese sentences*, PhD. University College London.
- Huddleston R. and Pullum G. (2002), *The Cambridge Grammar of the English Language*, Cambridge, Cambridge University Press.
- Hudson R. (1984), *Word Grammar*, Oxford, Blackwell.
- Hudson R. (1990), *English Word Grammar*, Oxford, Blackwell.
- Hudson R. (1994), Word Grammar, In Asher, R.(ed.), *Encyclopedia of Language and Linguistics*, Oxford, Pergamon, 4990-4993.
- Hudson R. (2003), Word Grammar, In Frawley, W.(ed.), *International Encyclopedia of Linguistics, Second Edition*, Oxford, Oxford University Press.
- Kahane S. (forthcoming), The Meaning-Text Theory, In Àgel,Vilmos; Eichinger,Ludwig; Eroms,Hans-Werner; Hellwig,Peter; Heringer,Hans-Jürgen; Lobin,Henning (eds.), *Dependency and Valency. An International Handbook of Contemporary Research*, Berlin, De Gruyter.
- Keenan E. (1976), Towards a universal definition of 'subject', In Li, C.(ed.), *Subject and Topic*, New York, Academic Press, 303-333.
- Melčuk I. (1988), *Dependency syntax : theory and practice*, Albany, State University Press of New York.
- Melčuk I. (1997), *Vers une Linguistique Sens-Texte*, Paris, Collège de France: Chaire Internationale.
- Rosch E. (1976), Classification of real-world objects: origins and representations in cognition, In Ehrlich, S. & Tulving, E.(eds.), *La Mémoire Sémantique*, Paris, Bulletin de Psychologie.
- Taylor J. (1995), *Linguistic Categorization: Prototypes in linguistic theory*, Oxford, Clarendon.
- Tomasello M. (2000), The item-based nature of children's early syntactic development, *Trends in Cognitive Sciences*, Vol. 4, 156-163.