Chains in Syntax and Morphology

Thomas Groß

Faculty of International Communication, Aichi University Machihata-cho 1-1, Toyohashi-shi, 441-8522 Aichi-ken, Japan tmgross@vega.aichi-u.ac.jp

Abstract. This paper argues for the existence of a deeper and more primitive structural unit of syntax and morphology than the constituent. Data from ellipsis, idiom formation, predicate complexes, bracketing paradoxes, and multiple auxiliary constructions challenge constituency-based analyses. In chain-based dependency grammar, however, constituents are seen as complete components. Components are units that are continuous both in the linear and in the dominance dimension. A unit continuous in the dominance dimension is called a chain. Evidence suggests that chains constitute the fundamental structural relationship between syntactic and morphological units, and that constituents are just a special subset of chains. If these assumptions are correct, linguistic research may need to change direction.

Keywords: bracketing paradox, chain, constituent, ellipsis, idiom

1 Introduction

In current linguistic research, the vast majority of work assumes that constituency is the ordering principle of syntax and morphology. Items that belong together form constituents. If items are understood as belonging together, but do not appear together, one assumes displacement of (at least) one item. This is then explained by movement (internal merge) in syntax, or by lowering in morphology.

Evidence from three areas in syntax, however, poses significant problems for a constituencybased analysis. These areas are ellipsis, idiom formation, and predicate verb complexes Examples of these phenomena are given in section 2.1.

In morphology, bracketing paradoxes in particular have shown considerable resistance to constituency-based analyses. Further, multiple auxiliary constructions challenge the constituency position. These areas are taken up in section 2.2.

This paper argues that languages need not necessarily build constituents, but rather chains.¹ The concept of the chain is introduced in section 3. In section 3.1, chains are applied to the syntactic phenomena introduced in section 2.1. Section 3.2 shows that chains equally apply in morphology. The bracketing paradox examples need to form proper chains, not proper constituents. Multiple auxiliary constructions also form chains, not constituents.

The paper concludes that, based on the evidence presented, the constituent may not be the central structural unit of syntax and morphology, but rather the chain is. If the aforesaid holds, research in syntax and morphology, but also in language acquisition and computer-based language modeling, should start to at least ponder the possible existence of a more central unit, and what this could mean for theory formation.

¹ To date, the system employed here has been referred to as "chain-based dependency grammar". Due to a number of criticisms, the contributors to this system have decided to change the terminology. "Chains" are now referred to as *catenae* (the Latin word for 'chains') in Osborne & Putnam (forthcoming). In this paper however, I will continue to use the older terminology. Any instance of "chain" in this paper may be substituted against "catena" in order to yield the new terminology.

2 Difficulties for Constituency-based Approaches

This section details some data that present evidence challenging the assumption that languages build their syntactic and morphological structures as constituents. The syntactic phenomena of ellipsis, idiom formation, and predicate verb complexes are illustrated in section 2.1. Section 2.2 takes up bracketing paradoxes, and presents multiple auxiliary constructions. Examples are taken from English, German, and Japanese.

2.1 Syntactic Issues

This section addresses syntactic phenomena challenging constituency-based analyses. The following phenomena will be illustrated: ellipsis, idioms, and predicate verb complexes.

Ellipsis comprises a variety of different phenomena: gapping, pseudo-gapping, V(P)-ellipsis, N-ellipsis, N'-deletion etc.²

(1) (2) (3)	I had preferred tea, and you had preferred coffee. I will want to paint the bicycle, and you will want to repair it. Er will rote Gummis essen, sie will gelbe Gummis essen .	gapping pseudo-gapping gapping & N-ellipsis
	he wants red jelly baby eat, she yellow	
(4)	'He wants to eat red jelly babies, she wants to eat yellow ones.' kare-wa otya-o non-de , kanozyo-wa koohii-o non-de i-ru.	backward pseudo-gapping
(4)	he-TOP green.tea-ACC drink-PRT, she-TOP coffee-ACC drink-PRT	1 011 0
	'He is drinking tea, she coffee.'	
(5)	Atasi, anata-o aisi-te i.ru. – Ore-mo omae-o-da.	V-ellipsis
	I[f] you-ACC love-PRT be-NPST. – I[m]-FOC you-ACC-COP	
	'I[f] love you. – I[m] do you, too.'	
(6)	omoi-das-u-ka, omoi- das-ana-i-ka-wa mondai-da.	LNR-gapping
	think-INC-NPST-INT,INC-NEG-NPST-INT-TOP problem-COP	Yatabe (2001: [5])
	'The problem is whether [you] remember or not.'	
(7)	aka-i Meizi-zidai-no tyawan-ga → aka-i-no-ga	N'-deletion
	red-NPST Meiji-period-GEN tea.bowl-NOM \rightarrow red-NPST-NR-NOM	[
	'[a] red tea bowl from the Meiji period \rightarrow [a] red one'	

The examples above present various instances of deletion, of which many are attributed to Ross (1967). In (1), the gap is *had preferred*. In (2), the gap is *want*. In German (3), the gap is *will Gummis essen*. For (1-3), derivational theories assume Right-Node-Raising. In Japanese (4), the gap is *nonde*. Since the gap precedes its antecedent, this phenomenon is called *backward* gapping. In Japanese (5), *aisite iru* is elided and replaced by the copula. In Japanese (6), the lexical part of a compound or derivation (depending on how *das-u* is categorized), does not appear in the second conjunct. Left-Node-Raising (LNR) has been proposed to explain this phenomenon (Yatabe 2001). In Japanese (7), *Meizi-zidai-no tyawan* is elided and replaced by *-no* in an operation called N'-deletion (Saito and Murasugi 1990).

Idiom formation does not always lead to constituents. O'Grady (1998) therefore argues that idioms are present in the lexicon as chains. In the next examples, the words constituting the idioms are italicized.

- (8) She was at her wits' end.
- (9) Sie war *mit* ihrem *Latein am Ende*. she was with her Latin at.the end 'She was at her wits' end.'

² Seminal contributions are Ross (1967), Jackendoff (1971), Hankamer (1979) etc. More recent contributions are from Benmamoun and Lappin (1999), Hartmann (2000), Johnson (2001), Merchant (2001), and from a dependency perspective Lobin (1993) and Osborne (2006, 2008).

- (10) People are *looking daggers at* me.
- (11) Hitobito-wa watasi-o *siro-i me-de mi-*te i-ru. people-TOP I-ACC white-NPST eye-INST see-PRT be-NPST

'People are looking daggers at me.'

In (8) and its German equivalent (9), the possessives are not part of the idioms. Therefore the idioms do not form constituents. In (10) and its Japanese equivalent (11), the object (of the preposition) is excluded from the idiom. Again, the words of the idioms do not form proper constituents.

Predicate verb complexes in continuous structures, i.e. without displacement, also do not form constituents, because they exclude other material. The verb complexes in (12-16) are italicized.

- (12) She had visited him.
- (13) She had been visiting him.
- (14) He had been visited by her.
- (15) Sie *hat* das nicht *machen wollen*. she has that not do want 'She has not wanted to do that.'
- (16) Kare-ni mot-te ki-te morat-te kudasai! he-DAT hold-PRT come-PRT receive-PRT give[respect].IMP 'Please let him bring [it]!'

In the English examples (12-14), the complex predicates form meaningful units to the exclusion of the object *him* in (12-13), and to the PP *by her* in (14). In German (15), the complex predicate *hat...machen wollen* does not form a constituent. In Japanese (16), the multiple participle construction also does not form a constituent. In order to form constituents, the predicate complexes must always include their objects (12, 13, 15, 16), PPs (14), and adverbs (15).

2.2 Morphological Issues

Bracketing paradoxes pose a recalcitrant problem for constituency-based analyses (Spencer 1988). The following illustration concentrates on two types of morphosyntactic bracketing paradoxes. The first type is known as morphosemantic mismatch (Beard 1991). The second type is a non-mismatching paradox.³ Expressions like *moral philosopher*, *theoretical linguist*, or *criminal lawyer* are instances of morphosemantic mismatching.

(17) a.*[moral] [philosoph-er]

b. [moral philosoph]-er]

The analysis (17a) groups words together. But because *moral* is understood as the attribute of the discipline philosophy, it must combine with *philosoph*, before the suffix *-er* attaches. Therefore (17b) should be the correct bracketing. (17b), however, contradicts assumptions about constituency, the fragment *moral philosoph* not being a viable constituent.

Non-mismatching bracketing paradoxes involve, for instance, non-peripheral plural formation in compounds.

- (18) many hangers-on
- (19) many aides-de-camp

³ The discussion neglects purely morphological bracketing paradoxes such as e.g. *uneasier*.

In (18-19), the plural morphemes are not in a peripheral position. The expressions *hanger-ons* and *aide-de-camps* are ungrammatical as plural forms. Although no semantic mismatch obtains, these expressions cannot be properly bracketed.

A problem of a similar sort occurs with multiple auxiliary constructions. The problem is that non-adjacent morphemes have a combined non-compositional function. The next example is from Anderson (1992: 16), based on Chomsky (1957: 39):

(20) John has been being seen.

Example (20) contains the perfective has+en, the progressive be+ing, and the passive be+n. In a lowering analysis, this would be rendered in the following manner:

(21) ... (has t_1) (be-en₁ t_2) (be-ing₂ t_3) see-n₃

The non-compositional meaning is established within the respective brackets. Constituenthood, however, is not achieved at the surface. Furthermore, it is unclear in which order the traces and the lowered suffixes have to be rendered.

The next section will show that deleted, idiomatic or predicate material need not form constituents, but must form chains, and that in morphology, morphemes need not form words, but must also form chains.

3 Chains

The chain is a unit of dependency grammar.⁴ Dependency grammar is primarily a word grammar, meaning that words form the building blocks of syntax, not constituents. Every word depends on exactly one other word, barring the root of the matrix clause, which does not depend on any other word. The dependency relationships to be established must conform to so-called constituency tests. Barring functional projections, properly constructed dependency trees can be interpreted in a fashion similar to phrase markers. The next example shows a dependency tree:



Barring the top node–and in many current theories–the assumption of a DP, (22) is largely identical to its phrase marker counterpart. The PP *of sentences*, for instance, is seen as a subtree in which the preposition *of* dominates the noun *sentences*. The angled edges are dependency edges, and the vertical edges are projection lines.⁵ Every word–barring the root node–must receive one of each of these edges.

Any two words that are directly connected by a dependency edge form a *chain*. Furthermore, every word node itself is considered as a chain. In (22), there are thus 34 distinct chains: *words*

⁴ Relevant contributions are: Tesnière (1959), Hays (1964), Robinson (1970), Kunze (1975), Matthews (1981, 2007), Mel'čuk (1988), Schubert (1988), Starosta (1988), Pickering & Barry (1993), Engel (1994), Jung (1995), Heringer (1996), Groß (1999), Eroms (2000), Kahane (2000), Tarvainen (2000), Ágel et al. (2003, 2006), and Osborne (2005).

⁵ Projectivity in dependency grammar is not understood as projection of nodes–as in mainstream constituency grammars–but as indicators of continuous sentence structure. Discontinuities show up as tangled edges, where a dependency edge crosses one or more projection lines.

themselves, words themselves form, words form, words themselves form structure, words form structure, words themselves form the structure, words form the structure of, words form structure of, words form the structure of, words form the structure of, words form the structure of sentences, words form the structure of sentences, words themselves form structure of sentences, words form structure of sentences, form the structure of sentences, the structure of, form the structure of, the structure of sentences, structure of sentences, the structure of, the structure of, the structure of sentences, structure of sentences, plus the individual words.⁶

A *string* is any combination of adjacent words. The words *words themselves* form a string, as do the words *themselves form*. There are 28 strings in (22), including the individual words.

A component obtains when the criteria for a chain and a string are both fulfilled. Words themselves is a component because it is both a chain and a string. themselves form, however, is not a component, because it is not a chain, but rather only a string. Words form is also not a component, because it is not a string, but rather only a chain. There are 21 components in (22): words themselves, words themselves form, words themselves form the structure, words themselves form the structure of, words themselves form the structure of sentences, form the structure of, the structure of, the structure of sentences, structure of sentences, plus the individual words.

If a component subsumes all dependents of all its nodes, then this component is *complete*. Complete components are *constituents*. There are 7 constituents in (22): *words themselves, themselves, the structure of sentences, of sentences, sentences, and the whole sentence.*

In summary, a chain is a word or a combination of words directly connected in the dominance dimension. A string is a word or a combination of words directly connected in the linear dimension. A component is both a string and a chain. A constituent is a complete component. The fact that "constituent" entails the notions of "chain", "string", and "completeness" indicates that constituents are neither primary nor primitive notions of syntax. Because every constituent is a component, and because every component is a chain, all constituents are chains.

The argument to be made in the rest of the paper is that the phenomena illustrated in section 2 are best analyzed in terms of chains.

3.1 Syntactic Chains

The discussion now returns to some examples in section 2.1. The gapping examples (1, 3) are given below as dependency trees, where the elided material in the second conjunct is italicized.

(i) Q: What do words themselves form? A: (Words themselves form) the structure of sentences.

Example (i) shows that the elision of the bracketed words results in an acceptable answer fragment. Hence, the chain *words themselves form* receives a justification, though maybe not an expected one. Also note that standard constituency tests will, of course, fail to identify chains, since chains are units without linear order properties, i.e. they are not strings. This, however, does not mean that chains are not compositional or not analyzable.

⁶ An anonymous reviewer contends that *words themselves form, words form structures* etc. "have no obvious justification". Due to space limitations it is impossible to provide a justification for all 34 chains in (22). The following brief discussion will limit itself to the first chain. The claim that chains exist as a primitive unit of syntax (and morphology) does not entail the claim that every chain must constitute a viable utterance. A chain can receive some sort of justification if its elision can render a remnant viable, as in an answer fragment. Consider (i):



In (23), the elided material *had preferred* is a chain, not a constituent. Equally in (24), the elided material *will...Gummis essen* forms a chain, not a constituent.

The idioms (8) and (11) are illustrated next. The words of the idioms are italicized.



In (25), the idiom *at...wits' end* forms a chain, but not a constituent. Equally, in (26), the idiom *siro-i me-de mi(-te)* is a chain, but not a constituent, because it is not complete as it excludes its object. It must be emphasized that the words forming idioms are always chains in the lexicon. In the syntax however, idiom chains are often broken up by syntactic processes.

The predicate complexes in (14-15) are also chains, but not constituents. In the dependency trees below the predicate chains are italicized.



The predicate complex *had been visited* in (27) is a chain, not a constituent. In (28), the predicate chain *hat...machen wollen* is a chain, not a constituent. The object *das* has risen and attached to a node dominating the governor of *das*, namely *machen*. The governor of *das* is marked by the subscripted 'G'. Rising is explained in detail in Groß & Osborne (2009).

In summary, elided material, idioms, and predicate complexes must form chains, but not necessarily constituents. These three areas comprise cognitively salient manifestations: If material to be deleted, and if idiom and predicate formation must correspond to chains, but not necessarily to constituents, then this is evidence for the existence of chains.

3.2 Morphological Chains

Extending chains into morphology avoids the problems of bracketing paradoxes and allows a parsimonious treatment of multiple auxiliary constructions. In the following, every morpheme is granted node status. Morphemes within and across words are connected in an equivalent fashion by dependencies. Because more nodes obtain in these trees, the edge conventions are as follows: Only the lowest morpheme of a word receives a projection edge, and intra-word dependency relationships are represented by dotted dependency edges. The next example (29) refers to (17b). Compare its structure to (30):



In (29-30), each word consists of two morphemes. In (29), the root noun *mor* is changed into an adjective by the attachment of *-al*. Therefore, *mor* depends on *-al*. Since *mor-al* is an attribute of philosophy, *mor-al* depends on the root noun *philosoph*. This root noun is turned into an agent noun by the attachment of *-er*. Therefore *philosoph* depends on *-er*, and as a result the chain *mor-al philosoph* depends on *-er*.

In (30), the noun *friend* is changed into an adjective by the attachment of -ly. Therefore, *friend* depends on -ly. This derived adjective functions as an attribute of an agent noun; therefore it cannot depend on *philosoph*, but must depend on *-er*. Note that the different dependencies of the attributive adjectives contribute to the respective but differing interpretations.

The dependency relationships in (29-30) correctly depict the semantic and morphological relationships between the constituent morphemes. Under a chain-based view, it is not necessary to first compile the words. Every chain can be viewed in isolation.

The chain-based solution to (18-19) is presented next:⁷



In (31), on depends on hang as it would in the verb phrase hang on. The attachment of *-er* transforms hang into an agent noun. Therefore, hang depends on *-er*. As a noun, *-er*, can take a plural form. In (31), *-er* depends on plural *-s*. A similar situation holds in (32). Note that the plural forms are outside the compounds in the dominance dimension. Bracketing cannot be

⁷ The expressions in (31-32) are treated here as compounds. Dependencies within compounds are depicted as syntactic, i.e. straight, dependency edges, but only the lowest node receives a projection edge. The assumption of a gradience from compounds to cliticization is possible, though. In the case of the latter, the straight dependency edges could be replaced with dotted edges reserved for morphological dependencies. Neither choice matters for the actual dependency structure, since it would remain identical.

properly applied, but a chain-based approach shows that the plural forms need not be peripheral in the linear dimension, but in the dominance dimension. Note the dependencies of *many*.

Morphological dependencies also help to explain N'-deletion in Japanese. The assumption of N'-deletion requires the adjective to reside in a specifier position. The adjective can, however, reside between two nominals as (33b) shows. (33a) repeats example (7).



Note that in (33ab), the italicized chain *Meizi-zidai-no tyawan* remains a proper chain, regardless of the position of *aka-i*. This chain is elided and replaced with *-no* in (34).



In conjunction with (33), (34) shows that what is called N'-deletion is actually not deletion of a constituent, as the name suggests, but deletion of a chain.

Multiple auxiliary constructions also receive a principled and parsimonious explanation. As in the previous examples, morphemes receive their own nodes as best possible. Consider again (20-21). A chain-based morphological dependency tree looks like (35):



The morphemes that together form functionally meaningful units also form chains. The perfective is expressed by the chain *has...-en*, the progressive is expressed by the chain *be...-ing*, and the passive is expressed by the chain *be...-n*. To the right, the pure dominance dimension is

shown. The wavy brackets indicate words. Note that the functional chains are immediate in the dominance dimension. Also note that word borders do not coincide with the boxed, functionally meaningful chains.

4 Conclusion

This paper has argued that the central and primitive unit of syntax and morphology is not the constituent, but the chain, a notion unique to dependency grammar. Evidence challenging constituency-based analyses has been provided and contrasted with a chain-based dependency analysis.

In the framework of this paper, constituents are complete components. The phenomena addressed here do not, however, require completeness, but rather a different feature of syntactic and morphological units, namely continuity in the dominance dimension. The unit that represents this notion to the exclusion of other features of constituents is the chain. When the frameworks that do not acknowledge the chain try to explain phenomena solely in terms of constituency, they must necessarily invoke all features of constituents. When certain phenomena cannot be satisfyingly explained using just the features of constituents, though, then some other tool is needed. The chain is just such a tool.

The chain is the key unit for accounts of ellipsis, idioms, predicate complexes and morphological structure. The evidence is strong for chains having linguistic and cognitive existence and relevance. The study of syntax and morphology should now question whether current constituency-based research is not headed in the wrong direction.

References

- Ágel, Vilmos, Ludwig Eichinger, Hans-Werner Eroms, Peter Hellwig, Hans Jürgen Heringer, and Henning Lobin eds. 2003. *Dependency and Valency: An International Handbook of Contemporary Research*, vol.1. Berlin: Walter de Gruyter.
- Ágel, Vilmos, Ludwig Eichinger, Hans-Werner Eroms, Peter Hellwig, Hans Jürgen Heringer, and Henning Lobin eds. 2006. *Dependency and Valency: An International Handbook of Contemporary Research*, vol.2. Berlin: Walter de Gruyter.

Anderson, Stephen R. 1992. A-Morphous Morphology. Cambridge: Cambridge University Press.

- Beard, Robert. 1991. Decompositional Composition: The Semantics of Scope Ambiguities and "Bracketing Paradoxes". *Natural Language and Linguistic Theory* 9, 195-229.
- Benmamoun, Elabbas and Shalom Lappin eds. 1999. *Fragments: Studies in Ellipsis and Gapping*, Oxford University Press, Oxford.
- Chomsky, Noam. 1957. Syntactic Structures. The Hague: Mouton & Co.
- Engel, Ulrich. 1994. Syntax der deutschen Gegenwartssprache, 3rd rev. ed. Berlin: Erich Schmidt.
- Eroms, Hans-Werner. 2000. Syntax der deutschen Sprache. Berlin: Walter de Gruyter.
- Groß, Thomas. 1999. Theoretical Foundations of Dependency Syntax. Munich: iudicium.
- Groß, Thomas and Timothy Osborne. 2009. Toward a Practical Dependency Grammar Theory of Discontinuities. *SKY Journal of Linguistics* 22, 43-90.
- Hankamer, Jorge. 1979. Deletion in coordinate structures. New York: Garland Publishing, Inc.
- Hartmann, Katharina. 2000. *Right Node Raising and gapping: Interface conditions on prosodic deletion*. Amsterdam: John Benjamins.
- Hays, David. 1964. Dependency theory: A Formalism and some Observations. *Language* 40, 511-525.
- Heringer, Hans Jürgen. 1996. Deutsche Syntax Dependentiell. Tübingen: Stauffenburg.
- Jackendoff, Ray. 1971. Gapping and related rules. Linguistic Inquiry 2, 21-35.

- Johnson, Kyle. 2001. What VP ellipsis can do, and what it can't, but not why. *The handbook of contemporary syntactic theory*. Mark Baltin and Chris Collins eds. Oxford: Blackwell Publishers, 439-479.
- Jung, Wha-Young. 1995. Syntaktische Relationen im Rahmen der Dependenzgrammatik. Hamburg: Buske.
- Kahane, Sylvaine (ed.). 2000. Les grammaires de dépendance. Traitement automatique des langues 41. Paris: Hermes.
- Kunze, Jürgen. 1975. Abhängigkeitsgrammatik. Studia Grammatica 12. Berlin: Akademie Verlag.
- Lobin, Henning. 1993. *Koordinationssyntax als prozedurales Phänomen*. Studien zur deutschen Grammatik 46. Tübingen: Narr.
- Matthews, Peter. 1981. Syntax. Cambridge University Press.
- Matthews, Peter. 2007. Syntactic Relations. Cambridge University Press.
- Mel'čuk, Igor. 1988. Dependency Syntax: Theory and Practice. Albany: State University of New York Press.
- Merchant, Jason. 2001. The syntax of silence: Sluicing, identity, and the theory of ellipsis. Oxford University Press.
- O'Grady, William. 1998. The Syntax of Idioms. *Natural Language and Linguistic Theory* 16, 79-312.
- Osborne, Timothy. 2005. Beyond the constituent: a dependency grammar analysis of chains. *Folia Linguistica* 39(3-4), 251-297.
- Osborne, Timothy. 2006. Parallel conjuncts. *Studia Linguistica* 60(1), 64-96.
- Osborne, Timothy. 2008. Major constituents: and two dependency grammar constraints on sharing in coordination. *Linguistics* 46(6), 1109-1165.
- Osborne, Timothy and Michael Putnam (forthcoming). Catenae: introducing a novel unit of syntactic analysis. *Syntax*.
- Pickering, Martin and Guy Barry. 1993. Dependency Categorial Grammar and coordination. *Linguistics* 31, 855-902.
- Robinson, Jane. 1970. Dependency structures and transformational rules. Language 46, 259-285.
- Ross, John R. 1967. Constraints on Variables in Syntax. Ph.D. dissertation, MIT.
- Saito, Mamoru and Keiko Murasugi. 1990. N'-Deletion in Japanese: A Preliminary Study. Japanese/Korean Linguistics 1, 285-301.
- Schubert, Klaus. 1988. *Metataxis: Contrastive Dependency Syntax for Machine Translation*. Foris, Dordrecht.
- Spencer, Andrew. 1988. Bracketing Paradoxes and the English Lexicon. Language 64, 663-682.
- Starosta, Stanley. 1988. *The Case for Lexicase: An Outline of Lexicase Grammatical Theory*. New York: Pinter Publishers.
- Tarvainen, Kalevi. 2000. *Einführung in die Dependenzgrammatik*. 2. Auflage. Tübingen: Niemeyer.
- Tesnière, Lucien. 1959. Éléments de syntaxe structurale. Klincksieck: Paris.
- Yatabe Shuichi. 2001. The syntax and semantics of left-node raising in Japanese. In Dan Flickinger and Andreas Kathol eds., *Proceedings of the 7th International Conference on Head-Driven Phrase Structure Grammar*, CSLI, Stanford, 325-344.