The minimalist program in syntax

Howard Lasnik

The Minimalist program, a development of earlier work in transformational generative grammar, proposes that the computational system central to human language is a 'perfect' solution to the task of relating sound and meaning. Recent research has investigated the complexities evident in earlier models and attempted to eliminate them, or to show how they are only apparent, following from deeper but simpler properties. Examples of this include the reduction of the number of linguistic levels of representation in the model, and the deduction of constraints on syntactic derivations from general considerations of economy and computational simplicity.

Minimalism is the latest development of TRANSFORMATIONAL GENERATIVE GRAMMAR (see Glossary), initiated by Noam Chomsky [1–4]. This approach to language centers on two psychological questions: (1) How is linguistic ability, (tacit) knowledge of language, represented in the human mind?; and (2) How does that knowledge arise in the individual? Chomsky presupposes a distinction between the knowledge ('competence') and how that knowledge is put to use in producing and understanding sentences ('performance'). In the realm of syntax, Chomsky's answer to the first question is that competence is a computational system whereby derivations produce structural representations. His answer to the second question (the question of 'explanatory adequacy' [5]) is that much of the computational system is innate, with, for the most part, only properties of particular lexical items having to be learned. The empirical bases for this answer are the deep similarities of the computational systems of languages, even unrelated ones, and the fact that speakers know far more about their languages than they would ever have had evidence for from the input (the 'poverty of the stimulus' argument).

The Minimalist program maintains that the derivations and representations constituting linguistic competence conform to an 'economy' criterion demanding that they be minimal in a sense determined by the language faculty (ultimately by general properties of organic systems): that is, there are no extra steps in derivations, no extra symbols in representations, and no representations beyond those that are conceptually necessary.

The immediate antecedents of the Minimalist program

There are several influential approaches to human language syntax, including Lexical-Functional Grammar (LFG) [6], Head Driven Phrase Structure Grammar (HPSG) [7], and Role and Reference Grammar (RRG) [8]. Minimalism developed out of the 'Government-Binding' (GB) or the PRINCIPLES and PARAMETERS model [9–11]. In that model, there are four significant levels of representation (see Fig. 1), related by derivation, as follows:

Items from the lexicon are inserted into the D-STRUCTURE in accordance with their syntactic properties and semantic roles, including thematic (θ) relations (agent of…, patient of…, etc., roughly corresponding in simple cases to subject of…, object of…). TRANSFORMATIONS successively alter the D-structure (the movement transformations leaving TRACES) eventually producing an S-STRUCTURE. For instance, in a passive sentence such as that in Fig. 2, the thematic object is transformationally displaced to subject position (and the auxiliary verb is raised to INFL; see Glossary), as in the D-structure and S-structure given in simplified form in Fig. 2a and 2b, respectively.

Transformations continue the derivation from S-structure to LF (in this instance producing no major changes). Phonological rules continue the derivation from S-structure to PF (with the traces deleted). Given that a human language is a way of relating sound (or, more generally, gesture, as in sign languages) and meaning, the interface levels PF and LF were assumed to be essential. Minimalism begins with the hypothesis that there are no other levels. Given traces, the role of D-structure in determining thematic relations becomes insignificant, as derived structure (augmented by traces) includes the relevant D-structure information.

Structure building: the old is new again

Minimalism, in a partial return to the apparatus of pre-1965 transformational theory [12], has lexical items inserted throughout the course of the syntactic derivation, via generalized transformations, rather than all in one initial block. The derivation proceeds 'bottom-up' with the most deeply embedded structural unit created, then combined with the head of which it is the complement to create a larger unit, and so on (see also Box 1). Consider the derivation of the sentence '*The woman will see the man*'. The noun (N) *man* is combined with the determiner (D) *the* to form the determiner phrase (DP) *the man*. This DP then combines with the verb *see* to produce an intermediate projection, V-bar. The DP *the woman* is created in the same fashion as *the man*, and is combined with the V-bar to produce the VP. Next, this VP merges with the Infl *will* producing I-bar. The DP *the woman* finally moves to the specifier position of I, yielding the full clausal projection IP, schematically illustrated below (by labeled bracketing, a notational variant of tree representation):

(1) $[_{IP}$ The woman $[_{I'}$ will $[_{VP}\ t\ [_{V'}$ see $[_{DP}$ the man]]]]]

In the more complex '*You think the woman will see the man*', the derivation of the embedded sentence is exactly as just outlined. The results of that derivation are combined with the verb *think*, forming a V-bar, which

Howard Lasnik
Dept of Linguistics,
University of Maryland,
1106 Marie Mount Hall,
College Park,
MD 20742, USA.
e-mail:
lasnik@wam.umd.edu

 1364-6613/02/$ – see front matter PII: S1364-6613(02)01977-0

Glossary

A(rgument)-position: subject position, direct object position, indirect object position, etc. in a sentence.
A-movement: movement to an A-position.
Anaphor: a linguistic expression entirely dependent on another linguistic expression (its antecedent) for its interpretation: e.g. *himself* in '*John injured himself*' is dependent upon *John*. In the GB framework, anaphors and their antecedents are **co-indexed** (annotated with the same numeric subscript).
Anaphora: the relation between an anaphor and its antecedent.
Binding: X binds Y iff X c-commands Y and X and Y are co-indexed. Binding theory concerns binding relations, and constraints on such relations.
Case filter: the requirement that in the course of a derivation, a nominal expression must eventually pass through or wind up in a position appropriate to its case. Such a position is called a position where the case is licensed.
Case positions: (see Table 1)

Table 1. Case positions

Position	Case	Example
Subject of finite sentence	Nominative	*He* left
Direct object of transitive verb	Accusative	I saw *him*
'Subject' of NP	Genitive	*John's* book
Object of preposition	'Oblique'	near *him*

Case theory: posits that the case distinctions (nominative, accusative, etc.) morphologically manifested in languages like Latin and Russian are present on nominal expressions in all languages. This more abstract notion of case is called **Case**.
Complementizer: a head that takes a clause (IP) as its complement, creating a Complementizer Phrase.
Covert syntax: the portion of a derivation between S-structure and LF (so-called because transformational operations here have no phonetic effects).
Cycle: under cyclicity, a domain of application of transformations and/or the sequence of transformations that applies in that domain.
Cyclicity: the requirement that transformations apply first on the most deeply embedded portion of a structure, then the next most deeply embedded, and so on.
D(eep)-structure: the starting point of a syntactic derivation in the 'standard theory' and in the GB theory. Grammatical relations ('subject of', 'object of', etc.) are structurally represented in D-structure.
Infl: the head containing tense information (e.g. past versus present) and agreement information (person, number, gender). Takes a VP as its complement to form an I(nfl)P (a clause).
Lexical insertion: inserting lexical items into a syntactic structure.
Locality constraints: constraints limiting how far, in hierarchical terms, a movement operation can displace an item. Most such constraints prohibit the movement of an item inside a certain structural configuration to a position outside of it. Also called 'isalnd constraints'.
Logical Form (LF): the syntactic structure in which all structural properties relevant to semantic interpretation are represented. The culmination of the syntactic derivation.
Multiple spell-out: the proposal that syntactic derivation interfaces with phonology and semantics throughout the derivation rather than at the two specific derivational points PF and LF.
Numeration: the selection of lexical items that will be used to construct a sentence (and an indication of how many times each will be used in that sentence).
Overt syntax: the portion of a derivation between D-structure and S-structure.
Parameter: a two (or more) valued choice determining a general property distinguishing one type of language from another.
Phonetic Form (PF): the syntactic structure in which all structural properties relevant to phonetic interpretation are represented lexical insertion.
Phrase marker: a representation of the abstract structural properties of a sentence.
Principle: a universal property of human language, assumed to be innate.
Reconstruction: phenomena where a transformationally moved item is interpreted in its pre-movement position.
S(urface)-structure: the phrase marker in a derivation which then branches towards PF and LF.
Strong feature: a property of a head requiring that something move to a position near it in overt syntax.
Trace: a marker left behind indicating the position from which transformational movement has taken place.
Transformation: an operation changing one phrase marker into another (a **singulary transformation**), or combining one with another (a **generalized transformation**). A series of such operations is a **derivation**, culminating (in the Government-Binding framework) in a **Logical Form** representation.
Transformational generative grammar: an approach to the syntax of human language, first developed by Noam Chomsky in the 1950s, based on phrase markers and transformations. 'Generative' means simply 'explicit' as in mathematics.
Weak feature: A property of a head requiring that something move to a position near it in covert syntax.
WH-movement: movement of an interrogative word or phrase to initial position in a clause. So-called because most of the interrogative words in English begin with *wh* (*who*, *what*, etc.)

then combines with *you*, and so on. In this model there is no one representation following all LEXICAL INSERTION and preceding all SINGULARY TRANSFORMATIONS: there is no D-structure.

Some Minimalist goals

So far, then, S-structure persists. If there is a point where the derivation divides, branching towards LF on one path and towards PF on the other, that point is S-structure. The more significant question is whether it has any of the further properties it has in the GB framework, for example, with respect to BINDING THEORY [9]. One goal of the Minimalist research program is to establish that these further properties are actually properties of LF (suggested in the mid 1980s [13], contrary to previous arguments [9]).

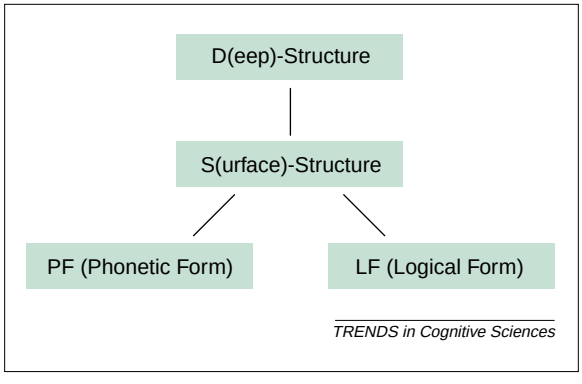

Fig. 1. Levels of representation in the Principles and Parameters model. (See text for details.)

Another goal is to reduce all constraints on representation to (presumably innate) bare output conditions, determined by the properties of the mental systems that LF and PF must interface with, for instance that a phonetic representation must be linearly ordered. Internal to the computational system, the desideratum is that constraints on transformational derivations be reduced to, again presumably innate, general principles of economy. Derivations beginning from the same NUMERATION are compared in terms of number of steps, length of movements, etc., with the less economical ones being rejected. An example is the Minimalist deduction of the 'Superiority Condition' [14], which demands that when multiple items are available for WH-MOVEMENT in a language, like English, allowing only one to move, it is the 'highest' one that will move:

(2) Who *t* will read what
(3) *What will who read *t* [* indicates ungrammaticality]

Economy, in the form of 'Shortest Move', selects (2) over (3) because the subject is closer to the

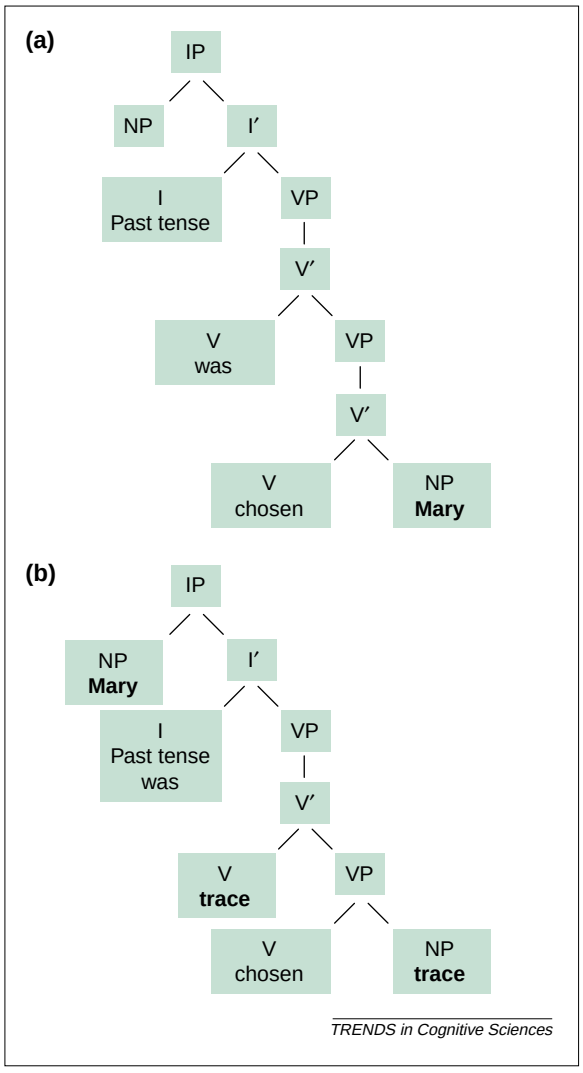

Fig. 2. The D-structure (a) and S-structure (b) of the sentence '*Mary was chosen*'. (See text for details.)

Box 1. Some terms of syntactic description

The following 'tree diagram' represents structured hierarchical organization of a phrase, XP, under X-bar theory*. It encodes the information that XP consists of YP and X' (X-bar), and X' consists of X and ZP. X is a head (an item taken from the lexicon); XP is a phrase based on (headed by) X, the maximal projection of X; and X' is an intermediate projection. YP and ZP are the maximal projections of Y and Z respectively. For simplicity, the internal structures of YP and ZP are suppressed.

```
      XP
     /  \
   YP    X'
        /  \
       X    ZP
```

YP is the **specifier** of X.
ZP is the **complement** of X.
YP and X' are **sisters**, as are X and ZP.
XP **dominates** YP, X', X, ZP, and everything that YP and ZP dominate.
XP **immediately dominates** YP and X'; X' immediately dominates X and ZP. [If A is the 'mother' of B, A immediately dominates B. Domination is the transitive closure of motherhood.]
A **c-commands** B if and only if every node dominating A also dominates B, and A does not dominate B. YP c-commands X', X, and ZP; X' c-commands YP.
A **asymmetrically c-commands** B if and only if A c-commands B and B does not c-command A. YP asymmetrically c-commands X and ZP.

***X-bar theory:** the proposal that the phrasal units of syntax are all based on ('projected from') a head (a minimal syntactic element, taken from the lexicon).

sentence-initial target of WH-movement than is the object.

The simplifying developments in the theory leading towards the Minimalist approach generally led to greater breadth and depth of understanding of both how human languages are organized ('descriptive adequacy') and how they develop in children's minds ('explanatory adequacy'). This success led Chomsky to put forward the audaciously Minimalist conjecture that the human language faculty might be a computationally perfect solution to the problem of relating sound and meaning, the minimal computational system given the boundary conditions provided by other modules of the mind. This conjecture leads to a general Minimalist critique of syntactic theorizing, including Chomsky's own earlier Minimalist theorizing. Consider first the leading idea that multiple derivations from the same initial set of lexical choices are compared. This introduces considerable complexity into the computation, especially as the number of alternative derivations multiplies. It thus becomes desirable to develop a model where all relevant derivational decisions can be made in strictly Markovian fashion: At each step, the very next successful step can be determined, and determined easily. This arguably more tractable 'local economy' model was suggested by Chomsky [15], developed in detail by Collins [16], then fully adopted by Chomsky [17]. Collins recently presented an overview of the approach [18].

The 'last resort' nature of syntactic movement

From its inception in the early 1990s, Minimalism has insisted on the 'last resort' nature of movement: Movement must happen for a formal reason. The CASE FILTER, which was a central component of the GB system, was thought to provide one such driving force. A standard example involves 'subject raising'.

(4) John is certain [*t* to fail the exam]
(5) It is certain [that John will fail the exam]

In (4), as in (5), *John* is the understood subject of *fail the exam*. This fact is captured by deriving (4) from an underlying structure much like that of (5), except with an infinitival embedded sentence instead of a finite one:

(6) __ is certain [John to fail the exam]

John in (6) is not in a position appropriate to any Case. By raising to the higher subject position, it can avoid a violation of the Case Filter, because the raised position is one where nominative case is licensed. But if the Case requirement of *John* provides the driving force for movement, the requirement will not be satisfied immediately upon the introduction of that nominal expression into the structure. Rather, satisfaction must wait until the next CYCLE, when a higher layer of structure is built, or, in fact, until an unlimited number of cycles later, as raising configurations can iterate:

(7) John seems [to be likely [to fail the exam]]

A Minimalist perspective favors an alternative where the driving force for movement can be satisfied immediately. Suppose that Infl has a feature that must be checked against the NP. Then as soon as that head

has been introduced into the structure, it 'attracts' the NP or DP that will check its feature. Movement is then seen from the point of view of the target rather than the moving item itself. The Case of the NP does get checked as a result of the movement, but that is simply a beneficial side effect of the satisfaction of the requirement of the attractor. In an elegant metaphor, Uriagereka likens the attractor to a virus [3]. Immediately upon its introduction into the body, it is dealt with (by the production of antibodies in the case of physical viruses, by movement to check the 'viral' feature in the syntactic case). The earlier Minimalist approach to the driving force of movement was called 'Greed' by Chomsky. This later one developed out of what I have called 'Enlightened Self Interest' [19].

The syntactic similarity of languages

One recurrent theme in GB and Minimalist theorizing, motivated by the quest for explanatory adequacy, is that human languages are syntactically very similar. The standard GB and early Minimalist instantiation of this claim was the proposal that superficial differences result from potential derivational 'timing' differences among languages, with the same transformation applying in OVERT or COVERT SYNTAX. Under both circumstances, LF reflects the results of the transformation. For example, the WH-movement operative in English interrogative sentences is overt movement to specifier of C(OMPLEMENTIZER). In many other languages, including Chinese and Japanese, interrogative expressions seem to remain '*in situ*', unmoved, as seen in the contrast between (8) and its English translation in (9).

(8) ni xihuan shei (you like who? – Chinese)
(9) Who do you like?

Huang [20] argued that even in such languages there is movement, by showing that well-established LOCALITY CONSTRAINTS on WH-movement [21] also constrain the distribution and interpretation of certain seemingly unmoved WH-expressions in Chinese. This argument was widely influential and laid the groundwork for much GB and Minimalist research. Along related lines, Chomsky argued that V-raising, overt in virtually all the Romance languages among others, operates covertly in English, as in the following examples from English and their translations into French:

(10) John often kisses Mary
*John kisses often Mary
(11) *Jean souvent embrasse Marie
Jean embrasse souvent Marie

The assumption is that the position of the verb vis-a-vis the adverb indicates whether the verb has raised overtly. For V-raising, the feature driving the movement is claimed to be one that resides in Infl. The feature might be STRONG (as in French) or WEAK. Similarly, the feature demanding overt WH-movement in English is a strong feature of C. The principle 'Procrastinate' disallows overt movement except when it is necessary (i.e. for the satisfaction of a strong feature [22,23]).

'Procrastinate' invited a question: why is delaying an operation until LF more economical than performing it earlier? Further, many of the hypothesized instances of covert movement do not have the semantic effects (with respect to quantifier scope, ANAPHORA, etc.) that corresponding overt movements have (see Ref. [4], Chapters 6 and 8). To address these questions, Chomsky in his most recent work [17,24] argues for a process of agreement (potentially at a substantial distance) that relates the two items that need to be checked against each other, similar to the HPSG position. Many of the phenomena that had been analyzed as involving covert movement are reanalyzed as involving no movement at all, just the operation Agree (although Huang's argument indicates that there is at least some covert movement). Overt phrasal movement (such as subject raising) is then seen in a different light: It is not driven by the need for Case or agreement features to be checked (as that could take place via Agree). Instead, it takes place to satisfy the requirement of certain heads (including Infl) that they have a specifier. Such a requirement was already formulated in the early GB framework [9], and dubbed the Extended Projection Principle (EPP) [10]. To the extent that long distance A-MOVEMENT as in (9) proceeds successive cyclically through each intermediate subject position, as is widely believed, the EPP is motivated, because, as observed above, these intermediate positions are not Case-checking positions. An important question at this point is why language has the seeming 'imperfection' of movement processes at all. Chomsky conjectures that phrasal movement is largely to convey topic-comment information (and possibly to make scope relations more transparent), and that the EPP is the way the computational system formally implements this. V-movement, on the other hand, is conjectured to have PF motivation (guaranteeing that the Infl affix will ultimately be attached to a proper host, V), and may even be PF process.

Syntactic interfaces

The connection between syntactic derivation and semantic and phonological interfaces has long been a central research area. In the earliest generative model [12], the interface is the T-marker, which includes all of the PHRASE MARKERS constituting the derivation. Subsequent models had the following interfaces with semantics:

(i) The 'standard theory' [5] → D-structure
(ii) Government-Binding (GB) → LF (via S-structure)
(iii) Early Minimalism → LF (via an uninterrupted transformational derivation beginning with the numeration)

The Minimalist approach to structure building is much like that of the 1950s [12], suggesting that interpretation in the Minimalist model also could be like that in the early model, distributed over many structures. Already in the late 1960s and early 1970s, Bresnan argued that the rule responsible for assigning English sentences their intonation contour applies following each cycle of

Questions for future research

- Is there any covert movement, or just 'Agree' (the process by which formal requirements can be satisfied at a distance)? If there is covert movement, what principles determine when it will take place?
- Under the copy theory of movement, what principles determine which parts of which copies are deleted, particularly in LF structures?
- How exactly does Multiple Spell-Out work? What are the precise principles determining how the sub-LF and PF structures created as successive phases are sent off for interpretation and combined into unified semantic and phonetic forms?
- For dealing with reconstruction effects, what is the proper division of labor between Multiple Spell-Out and the copy theory of traces?

transformations, rather than at the end of the syntactic derivation [25]. Jackendoff put forward similar proposals for semantic phenomena involving scope and anaphora [26]. In his recent Minimalist work [17,24], Chomsky argues for a general instantiation of this distributed approach (MULTIPLE SPELL-OUT), based on work of Epstein [27] and of Uriagereka [28]. At the end of each cycle (or 'phase' in Chomsky's most recent work), the syntactic structure thus far created can be encapsulated and sent off to the interface components for phonological and semantic interpretation. Thus, there are no levels of PF and LF. Epstein argues that such a move represents a conceptual simplification (in the same way that elimination of D-structure and S-structure does), and both Uriagereka and Chomsky provide some empirical justification. The role of syntactic derivation, always important in Chomskian theorizing, becomes even more central on this view. Epstein reasons that the centrality of (asymmetric) c-command (as opposed to one of a whole range of other conceivable geometric relations) in syntax is predicted on this strongly derivational view, but not in a more 'representational' theory. As the derivation proceeds, always merging together pairs of items, sisterhood and domination are the only immediately available primitives. And X (asymetrically) c-commands Y if and only if Y is dominated by the sister of X.

Multiple spell-out effectively deals with a range of RECONSTRUCTION phenomena. For example, an ANAPHOR normally requires an antecedent that c-commands it:

(12) John criticized himself

(13) *Himself criticized John

But when the anaphor is fronted from a position c-commanded by an antecedent to a position not in that structural relation, the anaphoric connection is nonetheless possible:

(14) Himself, John criticized

This follows straightforwardly if anaphora can be interpreted before movement.

Chomsky has also explored another kind of approach to reconstruction, based on a condition that he calls 'Inclusiveness' [15]. This condition demands that a syntactic derivation merely combine elements of the numeration. No new entities can be created. Traces, as traditionally conceived, violate this condition. Chomsky therefore concludes that a 'trace' of movement is actually a *copy* of the item that moved, rather than a new sort of entity. This is yet another return to earlier generative approaches (wherein movement was seen as a compound of copying and deletion). The copy left behind is normally deleted in the phonological component, (although Boskovic presents arguments that under certain circumstances lower copies are pronounced, to 'rescue' what would otherwise be PF violations [29]) but could persist for semantic purposes, such as the licensing of anaphoric connection. Fox presents an analysis of scope and anaphora reconstruction effects in terms of the copy theory [30].

An influential research line, initiated by Kayne [31], extends the impact of c-command to PF as well. Kayne hypothesizes that the linear order that is manifest in PF (as it must be, given properties of the phonetic system) comes about via his Linear Correspondence Axiom (LCA), which states that asymmetric c-command is mapped onto PF linear order. This has the far reaching consequence that structures must always be 'right-branching'. SVO languages like English are broadly consistent with this requirement, but SOV languages like Japanese are not. Kayne's 'antisymmetry' approach re-analyzes SOV languages as underlyingly SVO (as all languages must be by this hypothesis) with the SOV order derived by (leftwards) movement. One crucial unanswered question is the source of the 'driving force' for all of the required movements.

Conclusion

Chomsky constantly emphasizes that Minimalism is as yet still just an 'approach', a conjecture about how human language works ('perfectly'), and a general program for exploring and developing the conjecture. The explanatory success attained thus far gives some reason for optimism that the approach can be developed into an articulated theory of human linguistic ability and its development in the individual.

Acknowledgements
I would like to acknowledge the very helpful suggestions of Zeljko Boskovic, Noam Chomsky, and William Snyder, who commented on an earlier draft.

References

1 Chomsky, N. (1995) *The Minimalist Program*, MIT Press
2 Freidin, R. (1997) Review of Noam Chomsky, *The Minimalist Program*. *Language* 73, 571–582
3 Uriagereka, J. (1998) *Rhyme and Reason: An Introduction to Minimalist Syntax*, MIT Press
4 Lasnik, H. (1999) *Minimalist Analysis*, Blackwell
5 Chomsky, N. (1965) *Aspects of the Theory of Syntax*, MIT Press
6 Bresnan, J. (2001) *Lexical-Functional Syntax*, Blackwell
7 Pollard, C. and Sag, I. (1994) *Head Driven Phrase Structure Grammar*, CSLI
8 Van Valin, R. and LaPolla, R. (1997) *Syntax: Structure, Meaning, and Function*, Cambridge University Press
9 Chomsky, N. (1981) *Lectures on Government and Binding*, Foris
10 Chomsky, N. (1982) *Some Concepts and Consequences of the Theory of Government and Binding*, MIT Press
11 Chomsky, N. and Lasnik, H. (1993) The theory of principles and parameters. In *Syntax: An International Handbook of Contemporary Research* (Vol. 1) (Jacobs, J. *et al.*, eds), pp. 506–569, Walter de Gruyter
12 Chomsky, N. (1955) The logical structure of linguistic theory. Ms. Harvard University and MIT, Cambridge, MA, USA. (Revised 1956 version published in part by Plenum Press,1975; University of Chicago Press, 1985)

13 Chomsky, N. (1986) *Knowledge of Language*, Praeger
14 Chomsky, N. (1973) Conditions on transformations. In *A Festschrift for Morris Halle* (Anderson, S. and Kiparsky, P., eds), pp. 232–286, Holt, Rinehart & Winston
15 Chomsky, N. (1995) Categories and transformations. In *The Minimalist Program*, pp. 219–394, MIT Press
16 Collins, C. (1997) *Local Economy*, MIT Press
17 Chomsky, N. (2000) Minimalist inquiries: the framework. In *Step by Step: Essays on Minimalist Syntax in Honor of Howard Lasnik* (Martin, R. *et al.*, eds), pp. 89–155, MIT Press
18 Collins, C. (2001) Economy conditions in syntax. In *The Handbook of Contemporary Syntactic Theory* (Baltin, M. and Collins, C., eds), pp. 45–61, Oxford University Press
19 Lasnik, H. (1995) Case and expletives revisited: on Greed and other human failings. *Linguist. Inq.* 26, 615–633
20 Huang, C-T.J. ((1981) 82) Move *wh* in a language without *wh*-movement. *Linguist. Rev.* 1, 369–416
21 Ross, J.R. (1967) *Constraints on Variables in Syntax*, Doctoral dissertation MIT, Cambridge, MA, USA; published 1986 as *Infinite Syntax!*, Ablex
22 Chomsky, N. (1993) A minimalist program for linguistic theory. In *The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger* (Hale, K. and Keyser, S.J., eds), pp. 1–52, MIT Press
23 Lasnik, H. (1997) On feature strength: three minimalist approaches to overt movement. *Linguist. Inq.* 30, 197–217
24 Chomsky, N. (2001) Derivation by phase. In *Ken Hale: A Life in Language* (Kenstowicz, M., ed.), pp. 1–52, MIT Press
25 Bresnan, J.W. (1971) Sentence stress and syntactic transformations. *Language* 47, 257–281
26 Jackendoff, R. (1972) *Semantic Interpretation in Generative Grammar*, MIT Press
27 Epstein, S.D. (1999) Un-principled syntax: the derivation of syntactic relations. In *Working Minimalism* (Epstein, S.D. and Hornstein, N., eds), pp. 317–345, MIT Press
28 Uriagereka, J. (1999) Multiple spell-out. In *Working Minimalism* (Epstein, S.D. and Hornstein, N., eds), pp. 251–282, MIT Press
29 Boskovic, Z. (2001) *On the Nature of the Syntax–Phonology Interface: Cliticization and Related Phenomena*, Elsevier Science
30 Fox, D. (2000) *Economy and Semantic Interpretation*, MIT Press
31 Kayne, R.S. (1994) *The Antisymmetry of Syntax*, MIT Press

Connectionist natural language parsing

Dominic Palmer-Brown, Jonathan A. Tepper and Heather M. Powell

The key developments of two decades of connectionist parsing are reviewed. Connectionist parsers are assessed according to their ability to learn to represent syntactic structures from examples automatically, without being presented with symbolic grammar rules. This review also considers the extent to which connectionist parsers offer computational models of human sentence processing and provide plausible accounts of psycholinguistic data. In considering these issues, special attention is paid to the level of realism, the nature of the modularity, and the type of processing that is to be found in a wide range of parsers.

Dominic Palmer-Brown*
Leeds Metropolitan University, Computational Intelligence Research Group, School of Computing, Beckett Park, Leeds, UK, LS6 3QS.
*e-mail: d.palmer-brown@lmu.ac.uk

Jonathan A. Tepper
Heather M. Powell
The Nottingham Trent University, Department of Computing & Mathematics, Burton Street, Nottingham, UK NG1 4BU.
e-mail: jonathan.tepper@ntu.ac.uk; heather.powell@ntu.ac.uk

Connectionist parsers are neural-network-based systems (see Boxes 1 and 2) designed to process words or their syntactic types (tags) to produce a correct syntactic interpretation, or parse, of complete sentences. Parsers vary greatly in the way in which they tackle syntactic processing, and this is reflected in their modularity (or non-modularity) and in whether they combine neural networks with conventional symbolic processing to provide a hybrid solution, or adopt a purely connectionist approach.

Modularity and hybridity are reviewed as key attributes of connectionist parsers concerned with how the parsing problem is decomposed into (usually simpler) modules to form a parsing system, consisting of one or more connectionist modules and zero or more non-connectionist (e.g. symbolic) modules. The level of realism of parsers is assessed, by which we mean the ability of connectionist parsing systems to capture naturally occurring linguistic structures, behaviours and processing limitations. We discuss the extent to which parsers are able to capture the syntactic constraints and structures that naturally occur in language, as opposed to being limited to artificial grammars that restrict them to processing very small sub-domains of the language.

Modularity and hybridity

Psycholinguistic and fMRI-based evidence suggests that there is a significant component of purely syntactic processing of language that precedes and is independent of semantic processing [1]. This evidence can be taken as support for a syntactic module in a Fodorian sense [2], in that it is consistent with information encapsulation and spatial separation of processes. However, there is little consensus in terms of the details of how the syntactic module and its processing might be decomposed into separate modules performing distinct tasks. Within the connectionist research programme, many versions of modular architecture have been proposed, and as the neurocognitive evidence is still unclear, connectionists are free to explore the computational plausibility of different architectures. When a proposed architecture shows human-like performance on some aspect of syntax, it can be claimed as evidence of cognitive plausibility, but thus far no systems have been extended to a truly convincing range of language structures. What is clear, however, is that modular and hybrid parsers trained on corpora continue to make significant progress. Indeed, for large scale parsing, although there is little support for the eliminative connectionist viewpoint, which claims that purely connectionist systems are in principle capable of cognitive functions including language

http://tics.trends.com 1364-6613/02/$ – see front matter PII: S1364-6613(02)01980-0