15 The competence/performance distinction

The distinction between competence and performance (Chomsky 1965: Chapter I.1), which
is assumed by several theories of grammar, was already discussed in Section 12.6.3 about
the analysis of scrambling and verbal complexes in TAG. Theories of competence are in-
tended to describe linguistic knowledge and performance theories are assigned the task
of explaining how linguistic knowledge is used as well as why mistakes are made in
speech production and comprehension. A classic example in the competence/perfor-
mance discussion are cases of center self-embedding. Chomsky & Miller (1963: 286)
discuss the following example with recursively embedded relative clauses:

(1)   (the rat (the cat (the dog chased) killed) ate the malt)

(2b) is a corresponding example in German:

(2)   a. dass der Hund bellt, der   die Katze jagt,   die   die Maus kennt, die
          that the dog.M barks that.M the cat   chases that.F the mouse knows who
          im   Keller   lebt
          in.the basement lives
          'that the dog that chases the cat that knows the mouse who is living in the
          basement is barking'

      b. dass er Hund, [₁ der die Katze, [₂ die die Maus, [₃ die im   Keller
          that the dog   that the cat   that the mouse   who in.the basement
          lebt, ₃] kennt, ₂] jagt   ₁] bellt
          lives   knows   chases   barks

The examples in (1) and (2b) are entirely incomprehensible for most people. If one re-
arranges the material somewhat, it is possible to process the sentences and assign a
meaning to them.[1] For sentences such as (2b), it is often assumed that they fall within
our grammatical competence, that is, we possess the knowledge required to assign a

---

[1] The sentence in (2a) can be continued following the pattern that was used to create the sentence. For in-
stance by adding *die unter der Treppe lebte, die meine Freunde repariert haben* 'who lived under the staircase
which my friends repaired'. This shows that a restriction of the number of elements that depend on one
head to seven (Leiss 2003: 322) does not restrict the set of the sentences that are generated or licensed by
a grammar to be finite. There are at most two dependents of each head in (2a). The extraposition of the
relative clauses allows the hearer to group material into processable and reducible chunks, which reduces
the cognitive burden during processing.
    This means that the restriction to seven dependents does not cause a finitization of recursion ("Verend-
lichung von Rekursivität") as was claimed by Leiss (2003: 322). Leiss argued that Miller could not use his

structure to the sentence, although the processing of utterances such as (2b) exceeds language-independent abilities of our brain. In order to successfully process (2b), we would have to retain the first five noun phrases and corresponding hypotheses about the further progression of the sentence in our heads and could only begin to combine syntactic material when the verbs appear. Our brains become overwhelmed by this task. These problems do not arise when analyzing (2a) as it is possible to immediately begin to integrate the noun phrases into a larger unit.

Nevertheless, center self-embedding of relative clauses can also be constructed in such a way that our brains can handle them. Hans Uszkoreit gives the following example:

(3) Die Bänke, [₁ auf denen damals   die Alten   des   Dorfes, [₂ die allen
     the benches   on which back.then the old.people of.the village   that all
     Kindern, [₃ die vorbeikamen ₃], freundliche Blicke   zuwarfen ₂], lange Stunden
     children   that came.by   friendly   glances gave   long hours
     schweigend nebeneinander   saßen ₁], mussten im   letzten Jahr einem Parkplatz
     silent   next.to.each.other sat   must   in.the last   year a   car.park
     weichen.
     give.way.to

     'The benches on which the older residents of the village, who used to give friendly
     glances to all the children who came by, used to sit silently next to one another
     for hours had to give way to a car park last year.'

Therefore, one does not wish to include in the description of our grammatical knowledge that relative clauses are not allowed to be included inside each other as in (2b) as this would also rule out (3).

We can easily accept the fact that our brains are not able to process structures past a certain degree of complexity and also that corresponding utterances then become unacceptable. The contrast in the following examples is far more fascinating:[2]

(4) a. # The patient [who the nurse [who the clinic had hired] admitted] met Jack.
     b. * The patient who the nurse who the clinic had hired met Jack.

Although (4a) is syntactically well-formed and (4b) is not, Gibson & Thomas (1999) were able to show that (4b) is rated better by speakers than (4a). It does not occur to some people that an entire VP is missing. There are a number of explanations for this fact, all of which in some way make the claim that previously heard words are forgotten as soon as new words are heard and a particular degree of complexity is exceeded (Frazier 1985: 178; Gibson & Thomas 1999).

Instead of developing grammatical theories that treat (2b) and (4a) as unacceptable and (3) and (4b) as acceptable, descriptions have been developed that equally allow (2b),

---

insights regarding short term memory, since he worked within Transformational Grammar rather than in Dependency Grammar. The discussion shows that dependency plays an important role, but that linear order is also important for processing.

[2] See Gibson & Thomas (1999: 227). Frazier (1985: 178) attributes the discovery of this kind of sentences to Janet Fodor.

(3), and (4a) (competence models) and then additionally investigate the way utterances are processed in order to find out what kinds of structures our brains can handle and what kinds of structures it cannot. The result of this research is then a performance model (see Gibson (1998), for example). This does not rule out that there are language-specific differences affecting language processing. For example, Vasishth, Suckow, Lewis & Kern (2010) have shown that the effects that arise in center self-embedding structures in German are different from those that arise in the corresponding English cases such as (4): due to the frequent occurrence of verb-final structures in German, speakers of German were able to better store predictions about the anticipated verbs into their working memory (p. 558).

Theories in the framework of Categorial Grammar, GB, LFG, GPSG and HPSG are theories about our linguistic competence.[3] If we want to develop a grammatical theory that directly reflects our cognitive abilities, then there should also be a corresponding performance model to go with a particular competence model. In the following two sections, I will recount some arguments from Sag & Wasow (2011) in favor of constraint-based theories such as GPSG, LFG and HPSG.

15.1 The derivational theory of complexity

The first point discussed by Sag & Wasow (2011) is the Derivational Theory of Complexity. In the early days of Transformational Grammar, it was assumed that transformations were cognitively real, that is, it is possible to measure the consumption of resources that transformations have. A sentence that requires more transformations than the analysis of another sentence should therefore also be more difficult for humans to process. The corresponding theory was dubbed the *Derivational Theory of Complexity* (DTC) and initial experiments seemed to confirm it (Miller & McKean 1964; Savin & Perchonock 1965; Clifton & Odom 1966), so that in 1968 Chomsky still assumed that the Derivational Theory of Complexity was in fact correct (Chomsky 1976a: 249–250).[4] Some years later,

---

[3] For an approach where the parser is equated with UG, see Abney & Cole (1986: Section 3.4). For a performance-oriented variant of Minimalism, see Phillips (2003).

In Construction Grammar, the question of whether a distinction between competence and performance would be justified at all is controversially discussed (see Section 10.6.4.5.1). Fanselow, Schlesewsky, Cavar & Kliegl (1999) also suggest a model – albeit for different reasons – where grammatical properties considerably affect processing properties. The aforementioned authors work in the framework of Optimality Theory and show that the OT constraints that they assume can explain parsing preferences. OT is not a grammatical theory on its own but rather a meta theory. It is assumed that there is a component GEN that creates a set of candidates. A further component EVAL then chooses the most optimal candidate from this set of candidates. GEN contains a generative grammar of the kind that we have seen in this book. Normally, a GP/MP variant or also LFG is assumed as the base grammar. If one assumes a transformational theory, then one automatically has a problem with the Derivational Theory of Complexity that we will encounter in the following section. If one wishes to develop OT parsing models, then one has to make reference to representational variants of GB as the aforementioned authors seem to.

[4] In the Transformational Grammar literature, transformations were later viewed as a metaphor (Lohnstein 2014: 170, also in Chomsky 2001: Footnote 4), that is, it was no longer assumed to have psycholinguistic reality. In *Derivation by phase* and *On phases*, Chomsky refers once again to processing aspects such as computational and memory load (Chomsky 2001: 11, 12, 15; 2007: 3, 12; 2008: 138, 145, 146, 155). See also

however, most psycholinguists rejected the DTC. For discussion of several experiments that testify against the DTC, see Fodor, Bever & Garrett (1974: 320–328). One set of phenomena where the DTC makes incorrect predictions for respective analyses is that of elliptical constructions, for example (Fodor, Bever & Garrett 1974: 324): in elliptical constructions, particular parts of the utterance are left out or replaced by auxiliaries. In transformation-based approaches, it was assumed that (5b) is derived from (5a) by means of deletion of *swims* and (5c) is derived from (5b) by inserting *do*.

(5) a. John swims faster than Bob swims.
    b. John swims faster than Bob.
    c. John swims faster than Bob does.

The DTC predicts that (5b) should require more time to process than (5a), since the analysis of (5b) first requires to build up the structure in (5a) and then delete *swims*. This prediction was not confirmed.

Similarly, no difference could be identified for the pairs in (6) and (7) even though one of the sentences, given the relevant theoretical assumptions, requires more transformations for the derivation from a base structure (Fodor, Bever & Garrett 1974: 324).

(6) a. John phoned up the girl.
    b. John phoned the girl up.

(7) a. The bus driver was nervous after the wreck.
    b. The bus driver was fired after the wreck.

In (6), we are dealing with local reordering of the particle and the object. (7b) contains a passive clause that should be derived from an active clause under Transformational Grammar assumptions. If we compare this sentence with an equally long sentence with an adjective, like (7a), the passive clause should be more difficult to process. This is, however, not the case.

It is necessary to add two qualifications to Sag & Wasow's claims: if one has experimental data that show that the DTC makes incorrect predictions for a particular analysis, this does not necessarily mean that the DTC has been disproved. One could also try to find a different analysis for the phenomenon in question. For example, instead of a transformation that deletes material, one could assume empty elements for the analysis of elliptical structures that are inserted directly into the structure without deleting any material (see page 70 for the assumption of an empty nominal head in structures with

---

Marantz (2005: 440) and Richards (2015b).

A structure building operation that begins with words and is followed by transformations, as recently assumed by theories in the Minimalist Program, is psycholinguistically implausible for sentence parsing. See Labelle (2007) and Section 15.2 for more on incremental processing.

Chomsky (2007: 6) (written later than *On phases*) seems to adopt a constraint-based view. He writes that "a Merge-based system involves parallel operations" and compares the analysis of an utterance with a proof and explicitly mentions the competence/performance distinction.

noun ellipsis in German). Data such as (5) would then be irrelevant to the discussion.[5] However, reordering such as (6b) and the passive in (7b) are the kinds of phenomena that are typically explained using transformations.

The second qualification pertains to analyses for which there is a representational variant: it is often said that transformations are simply metaphors (Jackendoff 2000: 22–23; 2007: 5, 20): for example, we have seen that extractions with a transformational grammar yield structures that are similar to those assumed in HPSG. Figure 15.1 shows cyclic movement in GB theory compared to the corresponding HPSG analysis.
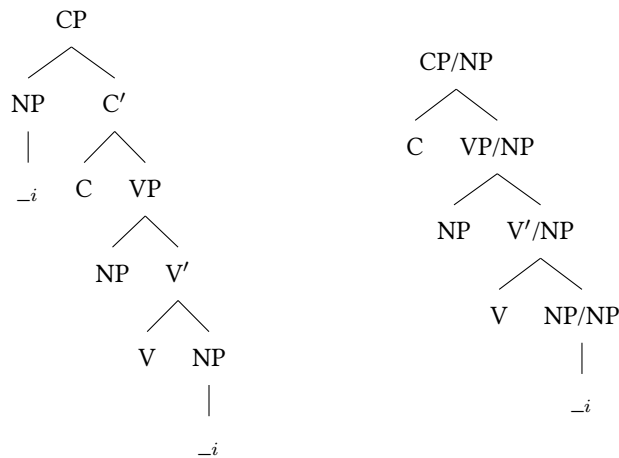

Figure 15.1: Cyclic movement vs. feature percolation

In GB, an element is moved to the specifier positions of CP (SpecCP) and can then be moved from there to the next higher SpecCP position.

(8) a. Chris$_i$, we think [$_{CP}$ _$_i$ Anna claims [$_{CP}$ _$_i$ that David saw _$_i$]].   (GB)
     b. Chris$_i$, we think [$_{CP/NP}$ Anna claims [$_{CP/NP}$ that David saw _$_i$]].   (HPSG)

In HPSG, the same effect is achieved by structure sharing. Information about a long-distance dependency is not located in the specifier node but rather in the mother node of the projection itself. In Section 19.2, I will discuss various ways of eliminating empty elements from grammars. If we apply these techniques to structures such as the GB structure in Figure 15.1, then we arrive at structures where information about missing elements is integrated into the mother node (CP) and the position in SpecCP is unfilled. This roughly corresponds to the HPSG structure in Figure 15.1.[6] It follows from this that there are classes of phenomena that can be spoken about in terms of transformations

---

[5] Culicover & Jackendoff (2005: Chapters 1 and 7) argue in favor of analyzing ellipsis as a semantic or pragmatic phenomenon rather than a syntactic one anyway.

[6] In Figure 15.1, additionally the unary branching of C′ to CP was omitted in the tree on the right so that C combines directly with VP/NP to form CP/NP.

without expecting empirical differences with regard to performance when compared to transformation-less approaches. However, it is important to note that we are dealing with an S-structure in the left-hand tree in Figure 15.1. As soon as one assumes that this is derived by moving constituents out of other structures, this equivalence of approaches disappears.

15.2 Incremental processing

The next important point mentioned by Sag & Wasow (2011) is the fact that both comprehension and production of language take places incrementally. As soon as we hear or read even the beginning of a word, we begin to assign meaning and to create structure. In the same way, we sometimes start talking before we have finished planning the entire utterance. This is shown by interruptions and self correction in spontaneous speech (Clark & Wasow 1998; Clark & Fox Tree 2002). When it comes to processing spoken speech, Tanenhaus et al. (1996) have shown that we access a word as soon as we have heard a part of it (also see Marslen-Wilson 1975). The authors of the study carried out an experiment where participants were instructed to pick up particular objects on a grid and reorganize them. Using eye-tracking measurements, Tanenhaus and colleagues could then show that the participants could identify the object in question earlier if the sound sequence at the beginning of the word was unambiguous than in cases where the initial sounds occurred in multiple words. An example for this is a configuration with a candle and candy: *candy* and *candle* both begin with *can* such that speakers could not yet decide upon hearing this sequence which lexical entry should be accessed. Therefore, there was a slight delay in accessing the lexical entry when compared to words where the objects in question did not contain the same segment at the start of the word (Tanenhaus et al. 1995: 1633).

If complex noun phrases were used in the instructions (*Touch the starred yellow square*), the participants' gaze fell on the object in question 250ms after it was unambiguously identifiable. This means that if there was only a single object with stars on it, then they looked at it after they heard *starred*. In cases where there were starred yellow blocks as well as squares, they looked at the square only after they had processed the word *square* (Tanenhaus et al. 1995: 1632). The planning and execution of a gaze lasts 200ms. From this, one can conclude that hearers combine words directly and as soon as enough information is available, they create sufficient structure in order to capture the (potential) meaning of an expression and react accordingly. This finding is incompatible with models that assume that one must have heard a complete noun phrase or even a complete utterance of even more complexity before it is possible to conclude anything about the meaning of a phrase/utterance. In particular, analyses in the Minimalist Program which assume that only entire phrases or so-called phases[7] are interpreted (Chomsky (1999) and Marantz (2005: 441), who explicitly contrasts the MP to Categorial Grammar) must

---

[7] Usually, only CP and vP are assumed to be phases.

therefore be rejected as inadequate from a psycholinguistic perspective.[8],[9]

With contrastive emphasis of individual adjectives in complex noun phrases (e. g. *the BIG blue triangle*), hearers assumed that there must be a corresponding counterpart to the reference object, e. g. a small blue triangle. The eye-tracking studies carried out by Tanenhaus et al. (1996) have shown that taking this kind of information into account results in objects being identified more quickly.

Similarly, Arnold et al. (2004) have shown, also using eye-tracking studies, that hearers tend to direct their gaze to previously unmentioned objects if the interlocutor interrupts their speech with *um* or *uh*. This can be traced back to the assumption that hearers assume that describing previously unmentioned objects is more complex than referring to objects already under discussion. The speaker can create more time for himself by using *um* or *uh*.

Examples such as those above constitute evidence for approaches that assume that when processing language, information from all available channels is used and that this information is also used as soon as it is available and not only after the structure of the entire utterance or complete phrase has been constructed. The results of experimental research therefore show that the hypothesis of a strictly modular organization of linguistic knowledge must be rejected. Proponents of this hypothesis assume that the output of one module constitutes the input of another without a given module having access to the inner states of another module or the processes taking place inside it. For example, the morphology module could provide the input for syntax and then this would be processed later by the semantic module. One kind of evidence for this kind of organization of linguistic knowledge that is often cited are so-called *garden path sentences* such as (9):

(9) a. The horse raced past the barn fell.

  b. The boat floated down the river sank.

The vast majority of English speakers struggle to process these sentences since their parser is led down a garden path as it builds up a complete structure for (10a) or (10b) only then to realize that there is another verb that cannot be integrated into this structure.

(10) a. The horse raced past the barn.

  b. The boat floated down the river.

---

[8] Sternefeld (2006: 729–730) points out that in theories in the Minimalist Program, the common assumption of uninterpretable features is entirely unjustified. Chomsky assumes that there are features that have to be deleted in the course of a derivation since they are only relevant for syntax. If they are not checked, the derivation crashes at the interface to semantics. It follows from this that NPs should not be interpretable under the assumptions of these theories since they contain a number of features that are irrelevant for the semantics and have to therefore be deleted (for an overview, see Richards 2015b). As we have seen, these kinds of theories are incompatible with the facts.

[9] It is sometimes claimed that current Minimalist theories are better suited to explain production (generation) than perception (parsing). But these models are as implausible for generation as they are for parsing. The reason is that it is assumed that there is a syntax component that generates structures that are then shipped to the interfaces. This is not what happens in generation though. Usually speakers know what they want to say (at least partly), that is, they start with semantics.

However, the actual structure of (9) contains a reduced relative clause (*raced past the barn* or *floated down the river*). That is the sentences in (9a) are semantically equivalent to the sentences in (11):

(11) a. The horse that was raced past the barn fell.
b. The boat that was floated down the river sank.

The failure of the parser in these cases was explained by assuming that syntactic processing such as constructing a sentence from NP and VP take place independently of the processing of other constraints. As Crain & Steedman (1985) and others have shown, yet there are data that make this explanation seem less plausible: if (9a) is uttered in a relevant context, the parser is not misled. In (12), there are multiple horses under discussion and each NP is clearly identified by a relative clause. The hearer is therefore prepared for a relative clause and can process the reduced relative clause without being led down the garden path, so to speak.

(12) The horse that they raced around the track held up fine. The horse that was raced down the road faltered a bit. And the horse raced past the barn fell.

By exchanging lexical material, it is also possible to modify (9a) in such way as to ensure that processing is unproblematic without having to add additional context. It is necessary to choose the material so that the interpretation of the noun as the subject of verb in the reduced relative clause is ruled out. Accordingly, *evidence* in (13) refers to an inanimate noun. It is therefore not a possible agent of *examined*. A hypothesis with *evidence* as the agent of *examined* is therefore never created when processing this sentence (Sag & Wasow 2011).

(13) The evidence examined by the judge turned out to be unreliable.

Since processing proceeds incrementally, it is sometimes assumed that realistic grammars should be obliged to immediately assign a constituent structure to previously heard material (Ades & Steedman 1982; Hausser 1992). Proponents of this view would assume a structure for the following sentence where every word forms a constituent with the preceding material:

(14) [[[[[[[[[[[[[[Das britische] Finanzministerium] stellt] dem] angeschlagenen]
the British treasury provides the crippled
Bankensystem] des] Landes] mindestens] 200] Milliarden] Pfund] zur] Verfügung].
banking.system of.the country at.least 200 billion pounds to use

'The British Treasury is making at least 200 billion pounds available to the crippled banking system.'

Pulman (1985), Stabler (1991) and Shieber & Johnson (1993: 301–308) have shown, however, that it is possible to build semantic structures incrementally, using the kind of phrase structure grammars we encountered in Chapter 2. This means that a partial semantic representation for the string *das britische* 'the British' can be computed without having to assume that the two words form a constituent in (14). Therefore, one

does not necessarily need a grammar that licenses the immediate combination of words directly. Furthermore, Shieber & Johnson (1993) point out that from a purely technical point of view, synchronous processing is more costly than asynchronous processing since synchronous processing requires additional mechanisms for synchronization whereas asynchronous processing processes information as soon as it becomes available (p. 297–298). Shieber and Johnson do not clarify whether this also applies to synchronous/asynchronous processing of syntactic and semantic information. See Shieber & Johnson (1993) for incremental processing and for a comparison of Steedman's Categorial Grammar and TAG.

What kind of conclusions can we draw from the data we have previously discussed? Are there further data that can help to determine the kinds of properties a theory of grammar should have in order to count as psycholinguistically plausible? Sag, Wasow & Bender (2003) and Sag & Wasow (2011, 2015) list the following properties that a performance-compatible competence grammar should have:[10]

- surface-oriented

- model-theoretic and therefore constraint-based

- sign-based organization

- strictly lexicalist

- representational underspecification of semantic information

Approaches such as CG, GPSG, LFG, HPSG, CxG and TAG are surface-oriented since they do not assume a base structure from which other structures are derived via transformations. Transformational approaches, however, require additional assumptions.[11] This will be briefly illustrated in what follows. In Section 3.1.5, we encountered the following analysis of English interrogatives:

(15)   [$_{CP}$ What$_i$ [$_{C'}$ will$_k$ [$_{IP}$ Ann [$_{I'}$ $\_k$ [$_{VP}$ read $\_i$]]]]].

This structure is derived from (16a) by two transformations (two applications of move α):

---

[10] Also, see Jackendoff (2007) for reflections on a performance model for a constraint-based, surface-oriented linguistic theory.

[11] An exception among transformational approaches is Phillips (2003). Phillips assumes that structures relevant for phenomena such as ellipsis, coordination and fronting are built up incrementally. These constituents are then reordered in later steps by transformations. For example, in the analysis of (i), the string *Wallace saw Gromit in* forms a constituent where *in* is dominated by a node with the label P(P). This node is then turned into a PP in a subsequent step (p. 43–44).

   (i)   Wallace saw Gromit in the kitchen.

While this approach is a transformation-based approach, the kind of transformation here is very idiosyncratic and incompatible with other variants of the theory. In particular, the modification of constituents contradicts the assumption of Structure Preservation when applying transformations as well as the *No Tampering Condition* of Chomsky (2008). Furthermore, the conditions under which an incomplete string such as *Wallace saw Gromit in* forms a constituent are not entirely clear.

(16) a. Ann will read what?
      b. * Will Ann read what

The first transformation creates the order in (16b) from (16a), and the second creates (15) from (16b).

When a hearer processes the sentence in (15), he begins to build structure as soon as he hears the first word. Transformations can, however, only be carried out when the entire utterance has been heard. One can, of course, assume that hearers process surface structures. However, since – as we have seen – they begin to access semantic knowledge early into an utterance, this begs the question of what we need a deep structure for at all.

In analyses such as those of (15), deep structure is superfluous since the relevant information can be reconstructed from the traces. Corresponding variants of GB have been proposed in the literature (see page 124). They are compatible with the requirement of being surface-oriented. Chomsky (1981a: 181; 1986a: 49) and Lasnik & Saito (1992: 59–60) propose analyses where traces can be deleted. In these analyses, the deep structure cannot be directly reconstructed from the surface structure and one requires transformations in order to relate the two. If we assume that transformations are applied 'online' during the analysis of utterances, then this would mean that the hearer would have to keep a structure derived from previously heard material as well as a list of possible transformations during processing in his working memory. In constraint-based grammars, entertaining hypotheses about potential upcoming transformation steps is not necessary since there is only a single surface structure that is processed directly. At present, it is still unclear whether it is actually possible to distinguish between these models empirically. But for Minimalist models with a large number of movements (see Figure 4.20 on page 150, for example), it should be clear that they are unrealistic since storage space is required to manage the hypotheses regarding such movements and we know that such short-term memory is very limited in humans.

Frazier & Clifton (1996: 27) assume that a transformation-based competence grammar yields a grammar with pre-compiled rules or rather templates that is then used for parsing. Therefore, theorems derived from UG are used for parsing and not axioms of UG directly. Johnson (1989) also suggests a parsing system that applies constraints from different sub-theories of GB as early as possible. This means that while he does assume the levels of representation D-Structure, S-Structure, LF and PF, he specifies the relevant constraints ($\overline{\text{X}}$ theory, Theta-Theory, Case Theory, …) as logical conditions that can be reorganized, then be evaluated in a different order, but logically equivalent, and be used for structure building.[12] Chomsky (2007: 6) also compares human parsing to working through a proof, where each step of the proof can be carried out in different orders. This view does not assume the psychological reality of levels of grammatical representation when processing language, but simply assumes that principles and structures play a role when it comes to language acquisition. As we have seen, the question of whether

---

[12] Stabler (1992: Section 15.7) also considers a constraint-based view, but arrives at the conclusion that parsing and other linguistic tasks should use the structural levels of the competence theory. This would again pose problems for the DTC.

we need UG to explain language acquisition was not yet decided in favor of UG-based approaches. Instead, all available evidence seems to point in the opposite direction. However, even if innate linguistic knowledge does exist, the question arises as to why one would want to represent this as several structures linked via transformations when it is clear that these do not play a role for humans (especially language learners) when processing language. Approaches that can represent this knowledge using fewer technical means, e.g. without transformations, therefore are preferable. For more on this point, see Kuhn (2007: 615).

The requirement for constraint-based grammars is supported by incremental processing and also by the ability to deduce what will follow from previously heard material. Stabler (1991) has pointed out that Steedman's argumentation with regard to incrementally processable grammars is incorrect, and instead argues for maintaining a modular view of grammar. Stabler has developed a constraint-based grammar where syntactic and semantic knowledge can be accessed at any time. He formulates both syntactic structures and the semantic representations attached to them as conjoined constraints and then presents a processing system that processes structures based on the availability of parts of syntactic and semantic knowledge. Stabler rejects models of performance that assume that one must first apply all syntactic constraints before the semantic ones can be applied. If one abandons this strict view of modularity, then we arrive at something like (17):

(17) $(\text{Syn}_1 \wedge \text{Syn}_2 \wedge \ldots \wedge \text{Syn}_n) \wedge (\text{Sem}_1 \wedge \text{Sem}_2 \wedge \ldots \wedge \text{Sem}_n)$

Syn_1–Syn_n stand for syntactic rules or constraints and Sem_1–Sem_n stand for semantic rules or constraints. If one so desires, the expressions in brackets can be referred to as modules. Since it is possible to randomly reorder conjoined expressions, one can imagine performance models that first apply some rules from the syntax module and then, when enough information is present, respective rules from the semantic module. The order of processing could therefore be as in (18), for example:

(18) $\text{Syn}_2 \wedge \text{Sem}_1 \wedge \text{Syn}_1 \wedge \ldots \wedge \text{Syn}_n \wedge \text{Sem}_2 \wedge \ldots \wedge \text{Sem}_n$

If one subscribes to this view of modularity, then theories such as HPSG or CxG also have a modular structure. In the representation assumed in the HPSG variant of Pollard & Sag (1987) and Sign-Based CxG (see Section 10.6.2), the value of SYN would correspond to the syntax module, the value of SEM to the semantic module and the value of PHON to the phonology module. If one were to remove the respective other parts of the lexical entries/dominance schemata, then one would be left with the part of the theory corresponding exactly to the level of representation in question.[13] Jackendoff (2000) argues

---

[13] In current theories in the Minimalist Program, an increasing amount of morphological, syntactic, semantic and information-structural information is being included in analyses (see Section 4.6.1). While there are suggestions for using feature-value pairs (Sauerland & Elbourne 2002: 290–291), a strict structuring of information as in GPSG, LFG, HPSG, CxG and variants of CG and TAG is not present. This means that there are the levels for syntax, Phonological Form and Logical Form, but the information relevant for these levels is an unstructured part of syntax, smeared all over syntactic trees.

for this form of modularity with the relevant interfaces between the modules for phonology, syntax, semantics and further modules from other areas of cognition. Exactly what there is to be gained from assuming these modules and how these could be proved empirically remains somewhat unclear to me. For skepticism with regard to the very concept of modules, see Jackendoff (2000: 22,27). For more on interfaces and modularization in theories such as LFG and HPSG, see Kuhn (2007).

Furthermore, Sag & Wasow (2015: 53–54) argue that listeners often leave semantic interpretation underspecified until enough information is present either in the utterance itself or the context. They do not commit to a certain reading early and run into garden paths or backtrack to other readings. This is modeled appropriately by theories that use a variant of underspecified semantics. For a concrete example of underspecification in semantics see Section 19.3.

In conclusion, we can say that surface-oriented, model-theoretic and strongly lexicalist grammatical theories such as CG, LFG, GPSG, HPSG, CxG and the corresponding GB/MP variants (paired with appropriate semantic representations) can plausibly be combined with processing models, while this is not the case for the overwhelming majority of GB/MP theories.