

Issues in the Choice of a Source for Natural Language Generation

David D. McDonald*
Brandeis University

The most vexing question in natural language generation is ‘what is the source’—what do speakers start from when they begin to compose an utterance? Theories of generation in the literature differ markedly in their assumptions. A few start with an unanalyzed body of numerical data (e.g. Bourbeau et al. 1990; Kukich 1988). Most start with the structured objects that are used by a particular reasoning system or simulator and are cast in that system’s representational formalism (e.g. Hovy 1990; Meteer 1992; Rösner 1988). A growing number of systems, largely focused on problems in machine translation or grammatical theory, take their input to be logical formulae based on lexical predicates (e.g. Wedekind 1988; Shieber et al. 1990).

The lack of a consistent answer to the question of the generator’s source has been at the heart of the problem of how to make research on generation intelligible and engaging for the rest of the computational linguistics community, and has complicated efforts to evaluate alternative treatments even for people in the field. Nevertheless, a source cannot be imposed by fiat. Differences in what information is assumed to be available, its relative decomposition when compared to the “packaging” available in the words or syntactic constructions of the language (*linguistic resources*), what amount and kinds of information are contained in the atomic units of the source, and what sorts of compositions and other larger scale organizations are possible—all these have an impact on what architectures are plausible for generation and what efficiencies they can achieve. Advances in the field often come precisely through insights into the representation of the source.

Language comprehension research does not have this problem—its source is a text. Differences in methodology govern where this text comes from (e.g., single sentence vs. discourse, sample sentences vs. corpus study, written vs. spoken), but these aside there is no question of what the comprehension process starts with.

Where comprehension “ends” is quite another matter. If we go back to some of the early comprehension systems, the end point of the process was an action, and there was linguistic processing at every stage (Winograd 1972). Some researchers, this author included, take the end point to be an elaboration of an already existing semantic model whereby some new individuals are added and new relations established between them and other individuals (Martin and Riesbeck 1986; McDonald 1992a). Today’s dominant paradigm, however, stemming perhaps from the predominance of research on question-answering and following the lead of theoretical linguistics, is to take the end point to be a logical form: an expression that codifies the information in the text at a fairly shallow level, e.g., a first-order formula with content words mapped to predicates with the same spelling, and with individuals represented by quantified variables or constants.

* 14 Brantwood Road, Arlington, MA 02174-8004; mcdonald@cs.brandeis.edu

It is somewhat puzzling that this question of where the comprehension process ends has apparently never been debated in the literature. Instead it seems largely taken for granted that the parsing process ends with the assembly of an expression in a suitable logic that captures the text's information content, perhaps with some functional annotations, and that a "reasoning" process then starts with that expression and draws inferences in order to resolve anaphors and establish the speaker's intent.

Problems arise when researchers project this default decomposition onto the process of producing language. All too often the process is divided into a "reasoning" and a "generation" component (see, e.g., Shieber, this issue)—an unfortunate choice of terminology because it reduces the scope of "generation" to triviality as we shall see. The primary motivation for the division is the desire for a bi-directional natural language processing system—one where the representation of the linguistic resources is *reversible* for use in both the comprehension and production of utterances. But while a reversible representation is indeed a proper goal for today's systems, the choice of logical form as the "pivot point" is problematic, especially a first-order formula.

A truly reversible linguistic mapping between intentional situation and utterance will have the comprehension process end where the generation process begins. Thus just as the psycholinguistically correct source for generation is still very much a matter of research (as it is even when the source is a computational object in a well-designed AI system), so too is the end-point of comprehension, and by implication the division of that process into components and representational levels. A declarative, reversible, form–meaning mapping does not ipso facto have to start/end at the level of logical form, but can originate at a much deeper level with the class definitions of the object types and relations of the speaker's conceptual model (McDonald 1993).

Considered in isolation, the production of text from a logical form is, quite frankly, trivial. It corresponds to the final "readout" phase of McDonald, Meteer, and Pustejovsky (1987), since all that remains to be done is to linearize its elements in keeping with the constraints of a surface grammar, carry out the trivial mapping to the phonetic (orthographic) forms of the words implicit in the predicates, and add the requisite grammatical function words and morphemes. This capability has been an established part of the state of the art for well over twenty years (see, e.g., Webber [1971], which is also the first work on reversible grammars for generation this author is aware of). Over the years new architectures for this "tactical" part of generation (we prefer the term *surface realization*) are introduced only because of new ideas in grammatical theory or in response to shifts in what is given in the immediately prior representational level.

In current research that focuses just on surface realization, all of the substantial tasks of generation are invariably subordinated to the "reasoner" or "strategic component," which is treated as a black box whose operations are seldom discussed. Examples of these tasks include construing the speaker's situation in realizable terms given the available vocabulary and syntactic resources (an especially important task when the source is raw data, e.g., precisely what points of the compass make the wind "easterly," [Bourbeau et al. 1990]); selecting the information to include in the utterance and deciding whether it should be stated explicitly or left for inference; distributing the information into sentences and giving it an organization that reflects the intended rhetorical force, coherence, and necessary cohesion given the prior discourse; and finding a mapping of the information to linguistic resources that is collectively *expressible* (i.e., has a surface realization; see Meteer 1992). How one chooses to approach these tasks has substantial implications for the kinds of structures that a surface realization process can sensibly be given as input and may not be taken for granted. As a consequence, any generation architecture that is proposed without including an articulation

of the early stages of the process is issuing a large promissory note that it may not be able to redeem.

Often the choice of a two-component process in comprehension (and by extension in generation) is based on the judgment that linguistic knowledge can and should be restricted to its own component, the one responsible for the form and content of grammatical rules, leaving to the other component all matters of general reasoning ("reasoners should not have to truck with grammatical issues"). This assumption has been seriously questioned within the generation community in recent years. The constraints imposed by the linguistic resources' limitations in what they are able to express and the delicacy of the conceptual and rhetorical choices state-of-the-art generators are called on to make combine to force a strong interdependency between early and late aspects of the process to the point where many generation researchers today do not recognize any strong division into components, with different aspects of linguistic knowledge appearing at many levels of representation (see Hovy, McDonald, and Young 1989).

All judgments about "components" are caught up in issues of modularity, information encapsulation, and the autonomy of syntax (see, e.g., Fodor 1983), issues that cannot be settled without substantial empirical experiment and theoretical argument. That notwithstanding, it already seems evident that if one incorporates within the purview of a "reasoner" such text planning activities as those listed earlier then it will be very hard to sustain the argument that knowledge of grammar can be restricted to just surface realization. Different aspects of this knowledge can still be relatively segregated, however; in particular it seems likely that early generation decisions only require tacit knowledge of what lexemes and constructions the language provides, without yet requiring access to phonetic forms, the assembly of detailed sequential structures, or the imposition of grammatical relations.

One of the more problematic aspects of taking the source for the generator to be a logical form is the very fact that it is represented as a single expression in a linear notation. This may seem a small matter of notation, but the computational properties of a logical form as it is usually represented give it a very low notational efficiency in generation (see Woods [1986] for a discussion of this notion). These include the simple fact that expressions must be scanned and parsed before the information they contain can be deployed, the lack of decompositional locality because of the use of scoping quantifiers and variables to represent individuals (Mellish 1985), and, indeed, as Shieber (this issue) points out, there is the question of the formula's intended structural realization, since the logical connectives that link a formula's terms underspecify their corresponding syntactic constructions because of the equivalence of other formulas under commutativity, associativity, and other truth-preserving logical transformations.

The force of much of Shieber's argument in regard to logical-form equivalence rests on the constraint imposed by bi-directional processing. If the choice of the information that an utterance is to express is made by a component with no knowledge of what the syntactic and lexical resources of the language are able to convey, then it is highly unlikely that its representation of the information will match what the comprehension process will arrive at as its representation of what the utterance meant—Shieber's notion of "canonical logical form."

Today's generators confront the problem regularly, as for example when a knowledge-based system passes just the symbol 'red-porsche' to the generator and its designer wants the phrase "*the red porsche*," or "*that car*," or "*the red one*" produced as is contextually appropriate. Practical generators invariably interpose a special purpose interface between the raw representation of the application they are speaking for and

their own general linguistic rules so as to compensate for the raw form's weaknesses or linguistically inappropriate organization—to 'match impedances' as it were.

Seen as a transducer from meaning representations to surface forms, a generator is driven by the terms and (formal) syntactic structure of its inputs. In the ideal bi-directional system this mapping would be deterministic and reversible, but in practice it is nondeterministic, with the generator adding information to a source representation that severely underspecifies its target utterance. Mismatch with the output of comprehension is inevitable since the parser in effect picks out a fully specified representation, reading into its form a correspondence with syntactico-lexical discriminations that the knowledge-based system cannot appreciate. In particular, the syntax of today's sources' logics provide little useful guidance about the form of the surface utterance, or, alternatively, if the syntax is carefully attended to, it imposes a straight-jacket on the space of possible target utterances and limits the possibilities for fluent phrasing or adapting to the discourse context—a perennial problem with the 'direct production' generators used with expert systems.

While a (very) long-term solution to this problem waits on a fundamental redesign of meaning representations that would bring them into alignment with the requirements of language, we can take steps in this direction now by improving the source notation: Dispense with connected expressions in favor of dealing independently with the terms that would have comprised it.¹

We know that in any interesting system the logical form that specifies an utterance's meaning will be composed dynamically as the needs of the situation dictate, rather than being taken from a preconstructed repository, since if this were not the case there would be no possibility for the creative use of language to accommodate new situations. Given this, one has to ask why the components of the representation of the meaning would ever need to be assembled into an expression rather than entered directly into an early linguistic level of representation as soon as the need for them is appreciated. What work in generation does a formula do *qua* formula that cannot be done by its elements individually given a suitable representation?

The extension of an abstract linguistic plan through the incremental addition of elements is in fact a standard technique in generation.² A good example is Jeff Conklin's GENARO system (Conklin 1983; Arbib, Conklin, and Hill 1987), which produced paragraph-length descriptions of pictures of houses. GENARO selected the information it would include using a procedure known as "iterative proposing," whereby it selected successive atomic units of information from its database (a KL-One network) in a sequence determined by their relative *salience* given the perspective of the picture. The units corresponded to individuals (e.g., houses, fences, colors), categorizations

1 Since there would no longer be any logical connectives (the "glue" in the expressions) to be rendered in different but logically equivalent ways in a text, this technique also has the advantage that it reduces the possibilities for mismatches between the way the speaker formulates information and a comprehension system will represent its analysis of the corresponding text to just the more interesting cases of mismatches in the lexical semantics, e.g., "owns 40% of Ajax Corp." vs. "has a 40% stake in Ajax Corp."

2 Many of the ideas about bi-directional grammars and generation were developed by Shieber and Doug Appelt at SRI, which makes it interesting to note here that in the original version of Appelt's KAMP generator, knowledge of the grammar was distributed throughout the system and acted locally in close coordination with the system's planning decisions, making it rather like the approach being described here (Appelt 1982; p. 112). Appelt later shifted to using Martin Kay's Functional Unification Grammar (Kay 1979) to increase modularity, perspicuity, and robustness to revisions in the plan, while at the same time retaining the temporal interleaving of planning and linguistic realization, i.e., at no one moment during the processing was there ever a full logical formula corresponding to the eventual utterance (Appelt 1985; p. 110). The use of a FUG also of course directly facilitates bi-directional applications (Appelt 1989).

and properties of individuals, and the relations among them, each unit contributing a referent or content word(s) to the utterance.

As each unit was selected, it was immediately incorporated into an abstract linguistic level of representation³ in the position that best reflected its salience relative to the units that were there already. Thus the order in which units were selected had a potentially dramatic impact on the form of the final utterance. Consider, for example, the NP *"a white two-story house,"* embedded in the context *"This is a picture of ____"* at the beginning of a description. Following the rough heuristic that the most salient properties of an object are positioned closest to the head when realized as adjectives, this NP is the result of GENARO selecting four semantic units in the following order:

- \$house1—the referent, and the source of *"a ____"* given that the house is being newly introduced into the discourse)
- house(\$house1) *"house"*
- two-story-building(\$house1) *"two story ____"*
- color(\$house1, \$white) *"white ____"*

The order of the units' selection follows their decreasing relative salience: the numbers in this instance were 2.0, 1.0, .56, and .20 respectively. Had the house or its appearance in the picture been different, say switching the relative salience of the two properties, then the order of selection and the resulting NP would reflect this: *"a two-story white house."* In different contexts, these units could have different realization, e.g., *"[it] is two stories high."*

If we were to attempt to rationally reconstruct GENARO's selections as a standard logical form, e.g.

$$\exists(x) \text{ house}(x) \ \& \ \text{two-story-building}(x) \ \& \ \text{color}(x, \text{white})$$

we would not only have to parse this linear notation and have to introduce some canonical structural correspondences by which to direct its surface realization, but we would have lost the salience information that gave GENARO its special sensitivity to the particulars of the picture it was describing, markedly degrading its fluency.

This example illustrates not only that semantic representations should explicitly record information about salience, but also that the pivot point for bi-directional processing can be moved much deeper than is usually considered. In GENARO and a goodly number of other generators we have rules for the selection of a set of minimal semantic units and their organization into a text as just described. On the parsing side we have the systems cited earlier, whose outputs are comparable units added to or embellishing an existing semantic model of essentially the same sort as this style of generator starts from. Given such architectures, the move to properly reversible rules awaits only a declarative statement of the few remaining parts of these systems where the mappings have been formulated procedurally—a project that is already well advanced (McDonald 1991, 1992b).

Returning finally to the question of what processes should be given the label "generation," we must be very careful to avoid reflexively identifying generation as

³ Today this level would correspond to Meteer's "Text Structure" (1992). At this level there is a commitment to constituency, lexical choices for heads, and the structural relations of head-arguments and matrix-adjuncts. The structure overall is unordered.

the obverse of parsing. After all, the determination of where a "parser" leaves off and some non-text directed process of "general inferencing" takes over is very much a question of how individual systems are designed. We also have evidence from state-of-the-art systems that an incommensurate amount of processing is presently being done in the two directions, and consequently any attempt to make components correspond is suspect.

Existing comprehension systems as a rule extract considerably less information from a text than a generator must appreciate in generating one. Examples include the reasons why a given word or syntactic construction is used rather than an alternative, what constitutes the style and rhetoric appropriate to a given genre and situation, or why information is clustered in one pattern of sentences rather than another. There seems to be no reason in principle why comprehension systems couldn't notice such things, though of course their conclusions would have to be indeterminate since they don't have access to all the information the speaker used. More likely the present state of affairs is simply reflective of the fact that the generation of quality text is a harder task than its comprehension.

My own answer to the question of 'how far back does generation go' is that it may be considered to start at the first point where a speaker must appeal to her knowledge of language as she begins the process of carrying out some action through the use of language. This classification is of course principally a mechanism for delimiting a field of research, but it does also suggest that the way we might best arrive at Shieber's "AI-complete" solution to the question of how semantic information should be represented is through a careful study of the needs of the generation process.

References

- Appelt, Doug (1982). "Planning natural-language utterances to satisfy multiple goals." SRI Technical Note 259, Menlo Park, CA.
- Appelt, Doug (1985). *Planning English sentences*. Cambridge University Press.
- Appelt, Doug (1989). "Bidirectional grammars and the design of natural language generation systems." In *Theoretical Issues in Natural Language Processing*, edited by Wilks, 199-205. Lawrence Erlbaum.
- Arbib, Michael; Conklin, Jeffery; and Hill, Jane (1987). *From Schema-Theory to Language*. Oxford University Press.
- Bourbeau, L.; Carcagno, D.; Goldberg, E.; Kittredge, R.; and Polguère, A. (1990). "Bilingual generation of weather forecasts in an operations environment." In *Proceedings, 15th International Conference on Computational Linguistics (COLING-90)*. 90-92.
- Conklin, E. Jeffery (1983). "Data-driven indelible planning of discourse generation using salience." Doctoral dissertation, University of Massachusetts, Amherst, MA. Technical report 83-13.
- Hovy, Eduard (1990) "Unresolved issues in paragraph planning." In *Current Research in Natural Language Generation*, edited by Dale, Mellish, and Zock. Academic Press.
- Hovy, Eduard; McDonald, David; and Young, Sheryl (1989). "Current issues in natural language generation: An overview of the AAAI Workshop on Text Planning and Realization." *AI Magazine*, 10(3), 27-29.
- Fodor, Jerry (1983). *The Modularity of Mind*. The MIT Press.
- Kay, Martin (1979). "Functional grammar." In *Proceedings, 5th Annual Meeting of the Berkeley Linguistics Society*. University of California, Berkeley, CA, 142-158.
- Kukich, Karen (1988). "Fluency in Natural Language Reports." In *Natural Language Generation Systems*, edited by McDonald and Bolç. Springer-Verlag.
- Martin, Charles, and Riesbeck, Chris (1986). "Uniform parsing and inference for learning." In *Proceedings, AAAI-86*. Philadelphia, PA. Morgan-Kaufmann.
- McDonald, David (1993). "Reversible NLP by deriving the grammars from the knowledge base." In *Reversible Grammar in Natural Language Processing*. Kluwer Academic Publishers.
- McDonald, David (1992a). "An efficient chart-based algorithm for partial-parsing of unrestricted texts." In *Proceedings, 3rd Conference on Applied Natural Language Processing (ACL)*. Trento, Italy, 193-200.
- McDonald, David (1992b). "Type-driven

- suppression of redundancy in the generation of inference-rich reports." In *Aspects of Automated Natural Language Generation*, (Springer Verlag Lecture Notes in AI, Number 587), edited by Dale, Hovy, Rosner, and Stock, 73–88. Springer-Verlag.
- McDonald, David; Meteer (Vaughan), Marie; and Pustejovsky, James (1987). "Factors contributing to efficiency in natural language generation." In *Natural Language Generation: Recent Advances in Artificial Intelligence, Psychology, and Linguistics*, edited by Kempen, 159–181. Kluwer Academic Publishers.
- Meteer, Marie (1992). *Expressibility and the Problem of Efficient Text Planning*. Pinter Publishers.
- Mellish, Chris (1985). *Computer Interpretation of Natural Language Descriptions*. John Wiley.
- Rösner, Deitmar (1988). "The generation system of the SEMSYN project: Towards a task-independent generator for German." In *Advances in Natural Language Generation*, edited by Zock and Sabah. Pinter Publishers.
- Shieber, Stuart; van Noord, Gertjan; Pereira, Fernando; and Moore, Robert (1990). "Semantic-head-driven generation." *Computational Linguistics*, 16(1), 30–42.
- Shieber, Stuart (1993). "The problem of logical-form equivalence." *Computational Linguistics*, 19(1), 179–190.
- Webber, Bonnie (1971). "The case for generation." In *Papers Presented at the Seminar in Mathematical Linguistics*, Volume XIII, edited by Woods. Aiken Computer Laboratory, Department of Linguistics, Harvard University, Cambridge, MA.
- Wedekind, Jürgen (1988). "Generation as structure driven derivation." In *Proceedings, 13th International Conference on Computational Linguistics (COLING-88)*. Budapest, Hungary, 732–737.
- Winograd, Terry (1972). *Understanding Natural Language*. Academic Press.
- Woods, William (1986). "Important issues in knowledge representation." In *Proceedings, IEEE*. 74(10).