

## Focusing on focus: a formalization

Yan Zuo

Letteren/GM/CLS

Postbus 90153

5000LE Tilburg

The Netherlands

yzuo@kub.nl

### Abstract

We present an operable definition of focus which is argued to be of a cognito-pragmatic nature and explore how it is determined in discourse in a formalized manner. For this purpose, a file card model of discourse model and knowledge store is introduced enabling the decomposition and formal representation of its determination process as a programmable algorithm (FDA). Interdisciplinary evidence from social and cognitive psychology is cited and the prospect of the integration of focus via FDA as a discourse-level construct into speech synthesis systems, in particular, concept-to-speech systems, is also briefly discussed.

### 1. Introduction

The present paper aims to propose a working definition of focus and thereupon explore how focus is determined in discourse; in doing so, it hopes to contribute to the potential integration of a focus module into speech synthesis systems, in particular, concept-to-speech ones. The motivation largely derives from the observation that focus, though recognized as 'the meeting point of linguistics and artificial intelligence' (Hajicova, 1987) carrying significant discourse information closely related to prosody generation, has nonetheless appeared evasive and intractable to formalization. Most current speech synthesis systems simply take focus as the point of departure in an *a priori* manner whilst few have looked into the issue of how focus occurs as it is, namely, how focus is determined (by the speaker presumably) in the discourse. We aim to redress this inadequacy by first defining focus as a cognito-pragmatic category, which then enables a formal and procedural characterization of focus determination process in discourse, captured as focus determination algorithm (FDA). The FDA to be proposed is largely based on human-human dialogue (though space consideration precludes the full presentation of data), but is believed to be applicable to human-computer interaction as well. The study is characterized by its interdisciplinary approach, combining insights and inputs from linguistics, neuroscience and social psychology.

### 2. Defining focus: a cognito-pragmatic category

The term focus has been used in various senses, at

least six of which can be identified, i.e., phonological (Pierrehumbert, 1980; Ladd, 1996), semantic (Jackendoff, 1972; Prince, 1985), syntactic (Rochemont, 1986), cognitive (Sanford & Garrod, 1981; Musseler et al., 1995), pragmatic (Halliday, 1967), and AI-focus (Grosz & Sidner, 1986)<sup>1</sup>. We argue that, first, these multiple uses of focus, though resulting in conceptual confusion, hint at the central status of the notion in core as well as peripheral linguistics. Second, focus as occurs in discourse is best captured by referring to both the interlocutors' cognitive computation and constant interaction, in accordance with the dual (i.e., cognitive and pragmatic) nature of discourse per se (Nuyts, 1992). Of the six above-mentioned senses, the cognitive and pragmatic ones serve as the basis for the present definition, although the caveat is immediately made that the two aspects are to be fully integrated rather than merely added together. Moreover, neither is to be adopted blindly given certain shortcomings of previous accounts of each, such as a general vagueness militating against their effective application in speech technology.

In this connection, we define focus as a cognito-pragmatic category, calling for the introduction of the cognitive construct of discourse model in relation to knowledge store. Presumably, every typical adult communicator has at his/her disposal a vast and extensive knowledge store relating to the scenes and events occurring in the world he/she is in. The contents of the store are acquired via direct perception of the environment and, less directly, communication with others or reflection upon past acquisitions. Discourse entails the employment and deployment of the knowledge store, but in a specific discourse only a subset of it deemed relevant to the on-going discourse is incurred, given the economy principle of human cognitive system (Wilkes, 1997). We refer to this subset of knowledge store (KS) in operation for and in a given discourse as discourse model (DM) and hold it as bearing directly on focus. Following Levelt (1989:114), DM is 'a *speaker's* record of what he *believes* to be *shared* knowledge about the content of the discourse as it evolved' (my italics). Thus, it is a cognitive construct incorporating an interactive dimension of speaker-hearer mutual assessment; it is also an ongoing, dynamic one being constantly

<sup>1</sup> Though it needs to be cautioned that such a division into these six senses is more an analytic expedient than implying there is clear-cut boundaries between them.

updated as discourse progresses. Similarly, the DM and the KS are related in a dynamic way allowing for potentially constant, on-line interaction during the discourse which we refer to as 'dynamic inclusion'. This implies that when 'off-line' (i.e., when no discourse is actively going on), DM is included in KS, as indicated in Figure 1 below. By comparison, when 'on-line' (i.e., when participants are engaged in a discourse), the dynamic dimension becomes evident in both their inter-relation and the internal structuring of DM, as illustrated in Figure 2.

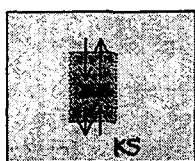


Figure 1. 'Off-line' state of DM in relation to KS

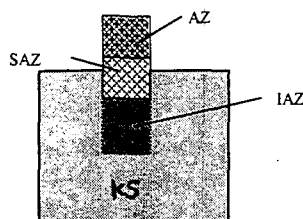


Figure 2. 'On-line' state of DM in relation to KS; AZ, SAZ & IAZ

Figure 2 deserves more explanation as the on-line state of and potential operations on the DM serve as the basis for focus determination in actual discourse. We argue that DM is crucially structured internally and for its representation we adopt the file card model based on the file metaphor in Heim (1983) (cf. also Reinhart, 1981; Vallduvi, 1992; Erteschik-Shir, 1997). A DM consists of a stack of file cards, and each card contains (maximally) three categories of items, viz., discourse referent (serving as index to and address of the card), attribute(s) and link(s), the first being obligatory whilst the latter two optional. Moreover, a card has one and only one referent but may have none, one or more attributes and links. Borrowing the notion of activation from Chafe (1987), we distinguish three zones, i.e., activated zone (AZ), semi-activated zone (SAZ) and inactivated zone (IAZ), within the DM<sup>2</sup>. Similar to the case of the DM-KS relation, the boundaries between the three zones are fluid rather than fixed, as is evident in Figure 2.

Armed with these machinery, we thus define focus as 'whatever is in the activated zone (AZ)', or, more precisely, whatever is at the top of the stack in AZ of the (speaker's version of the hearer's) DM as a result of immediately recent operations such as retrieval and updating at a given moment in the discourse (Zuo, 1999).

### 3. Focus determination algorithm (FDA)

Apparently, this definition of focus also renders the process of focus determination fairly transparent. The postulation of DM and KS enables the decomposition and characterization of the focus determination process in an explicit and formalized manner. Discourse is thereby reducible, to a considerable extent, to the operations on the file cards, most

crucially, adding, updating, locating and relocating of the cards across the three zones. In this vein, a card that is newly *added* to AZ (note not what *is* in AZ), or an item that is newly *entered* onto a card already in AZ at a specific moment is assigned focus-hood, *if and only if* the time interval between current moment and the moment for the addition/entry is shorter than a time threshold set on independent cognitive grounds (see below for more discussion). This process of focus determination can be represented as the following algorithm.

#### Focus Determining Algorithm (FDA)

```

1 SET 'file card in AZ (for the hearer)' (AZ (h)) = null
2 INPUT (message unit)
3 DO
4   Evaluator
5   Card Manager
6 INPUT (message unit)
7 UNTIL message unit = ender
8 END

Evaluator
9 EXTRACT discourse referent (Ri), attribute (Ai), and/or
  link (Li) from (the incoming) message unit
10 CREATE file card (Ci) indexed by Ri
11 COMPARE (Ci (= Ri (+ Ai) (+ Li)), {CAZ})
12 IF Ci ∈ {CAZ}
13 THEN
14   IF Ci ∈ {CSAZ} ∪ {CIAZ}
15   THEN
16     ADD Ci to AZ
17     RECORD time for addition Ta
18     LABEL Ci (with its content: Ri, (Ai), (Li)) FOCUS
19   ELSE
20     RETRIEVE file card indexed by Ri (Ci') from
      {CSAZ} ∪ {CIAZ}
21     ADD Ci' to AZ
22     RECORD time for retrieval Tr
23     LABEL Ci' (with its content: Ri', (Ai'), (Li')) FOCUS
24   ELSE
25     IDENTIFY Ci" in {CAZ} indexed by Ri
26     COMPARE (Ai, attribute(s) already on Ci" (Ai"))
27     IF Ai <> Ai"
28     THEN
29       ADD Ai to Ci"
30       RECORD time for addition Ta
31       LABEL Ai FOCUS
32     ELSE
33       COMPARE (Li, link(s) already present on Ci" (Li"))
34       IF Li <> Li"
35       THEN
36         ADD Li to Ci"
37         RECORD time Ta
38         LABEL Li FOCUS

Card Manager
39 SET Critical Time Threshold = Tt
40 RECORD Current Time = Tc
41 IF file card C ∈ {CAZ} at Tc AND Tc - Tr > Tt OR Tc - Ta > Tt
42 THEN
43   DEPOSIT C in IAZ
44 ELSE
45   IF C ∈ {CAZ} at Tc AND Tc - Tr ≅ Tt
46   THEN
47     DEPOSIT C in SAZ

```

Several notes are called for<sup>3</sup>. First, what can be

<sup>2</sup> Again here we are aware of the argument that activation is a continuous rather than a discrete concept.

<sup>3</sup> Due to space limit we only discuss a few major points here; for an elaborate account of the algorithm, refer to Zuo (1999).

assigned focus-hood? Obviously a slick (and vague) 'idea or thought' misses the point here. A look at the internal organization of the DM again suggests the answer. Corresponding to the content of the file card, four cases can be identified as to what can become the focus: (1) the discourse referent, (2) the attribute, (3) the link, and (4) the card as a whole. Note that this breakdown analysis meshes well with findings in psycholinguistic researches, for example, the possible candidates for acquiring 'conceptual prominence' distinguished in Levelt (1989:151). The file card model offers a more rigorous and operable way to account for such cases: Lines 16-18 and 20-23 respectively capture the above-mentioned cases (1) and (4) (though the former is apparently also a special type of case (4)) whilst Lines 29-31 and 36-38 respectively represent cases (2) and (3). Note that lines 16-18 and 20-23 show that a card may be added to AZ (and hence assigned focus-hood) either *ad externo* or by retrieving from SAZ or IAZ of the current DM.

Second, a crucial assumption of this algorithm is that speech planning consists of conceptual planning and linguistic planning proceeding in a sequential fashion; this is a well-established argument in psycholinguistics (Garrett, 1980), and the former proceeds in a unit-by-unit fashion (though the picture is more complicated for the latter) (Taylor & Taylor, 1990). Hence, the 'message unit' used in this algorithm (see Lines 2, 6, and 9) refers to such planning unit and can be roughly understood as 'chunk of meaning'; as such it consists minimally of a referent and an attribute while the link is optional; The 'ender' in Line 2 refers to the message unit intended by the speaker to terminate his/her current contribution. Obviously, here the speaker's intention plays a vital role. Note that the ender is also a conceptual unit in nature, and we leave open the question whether such enders constitute a closed, limited set with a relatively small number of prototypical units.

Third, the formula  $C_i = R_i (+A_i) (+L_i)$  in Line 11 indicates the make-up of the card, with the brackets standing for optionality (see Section 2). Also in this line, the function COMPARE ( $a$ ,  $b$ ) is defined as COMPARE  $a$  AGAINST  $b$ .  $\{C_{AZ}\}$  (and  $\{C_{SAZ}\}$ ,  $\{C_{IAZ}\}$  in the remainder of the algorithm) stands for the set comprised by the file cards already in AZ (or SAZ and IAZ, for that matter) at the current moment.

Fourth,  $T_a$  (L17),  $T_r$  (Ls 22, 36) and  $T_c$  (L39) refer to a point in time, in comparison with  $T_t$  (L38) which is an interval of time. They serve as input to the Card Manager sub-program which keeps track of the 'transportation', i.e., retrieval and deposition, of the cards. Thus, the RECORD (time) function (Ls 17, 22, 30, and 37), together with the Card Manager, takes care of the on-line shuffling and reshuffling of the file cards and is mainly responsible for the dynamism of DM. Regarding the choice of the threshold time  $T_t$  (L39), we argue that it is presumably the critical time conditioned by the capacity of the working memory;

but we leave open its specific value and on what terms, absolute or relative, it should be defined (for different views, cf. Carpenter, 1988; Liebert, 1997; Givon, 1983; Barbosa & Bailly, 1994). At present, the commonly-employed practice (which is also that adopted here) is to set a time threshold in terms of the length of some independently delimited discourse segments (e.g. those in Rhetoric Structure Theory (Hirscheberg, 1993)). We admit this inadequacy and wish to address it fully with inputs from interdisciplinary researches in the future.

Finally, the AZ, SAZ and IAZ in the algorithm refer to the hearer's DM as assessed by the speaker in discourse, i.e., the speaker's version of the hearer's DM, as the hearer's true DM is only accessible to s/he her/himself.

#### 4. Evidence from social and cognitive psychology

Crucially, the validity of FDA is contingent on (i) to what extent it is possible for the speaker to conceptualize the hearer's DM and (ii) on what independent grounds is the tripartite division of the DM justified? For the former question we invoke the notion of intersubjectivity from social psychology and for the latter, research findings in cognitive psychology are cited.

Stemming initially from the observation in social psychology that discourse participants have to constantly 'put themselves in each other's shoes' in order to achieve communicative goals (cf. Rommetveit, 1974; Clark, 1985), intersubjectivity is primarily concerned with perspective-taking, or, perspectivization (Sanders & Spooren 1997). It implies that discourse is a negotiating process and that understanding in discourse has to be sufficiently intersubjective. Hence, it is both necessary and possible for the speaker to assess the hearer's DM, and this is achieved through intersubjectivity. Admittedly, this process is not infallible, given Linell's (1995) observation regarding misunderstanding in discourse; nonetheless, it can be carried out with relative sufficiency which primarily depends on the participants' communicative competence and their expectation of the discourse.

A theory of discourse processing must also be a theory of cognition and memory; this is especially true for focus, given its attested relevance to memory. Researches on knowledge storage and processing in human memory in cognitive psychology have favored a dual memory system, i.e. working memory (WM) and long-term memory (LTM) (Baddeley, 1990) and a tripartite taxonomy of LTM into procedural, semantic, and episodic storage systems (Tulving, 1985). Moreover, WM serves as a portal to early episodic memory, and both are characterized by a limited capacity and rapid decay: the content in WM is periodically emptied into first, early episodic memory, then long-term episodic memory system, and thereafter semantic memory system. (e.g. Gathercole & Baddeley, 1993).

This representation dovetails nicely with our present account of focus and FDA. Specifically, a rough parallel may be drawn between, first, WM and AZ, second, early episodic memory and SAZ & IAZ, third, long-term episodic memory & semantic memory and IAZ & KS, and fourth, the dynamic working of knowledge processing and that of FDA, in particular the Card Manager which takes charge of the make-up of DM by constantly monitoring the timing and subsequently shuffling and reshuffling cards.

### 5. Integration of a focus module into speech synthesis systems

FDA, presented here on the basis of an operable definition of focus, enables the integration of a focus module into speech synthesis system; specifically, the output of FDA, i.e., the focus pattern of the message conveyed by the utterance, may be fed into a subsequent accent assignment module, one in the spirit of the Focus-Accent Theory of Dirksen (1992) and Dirksen & Quene (1993).

In this way, FDA entertains a great potential for the integration of discourse-level information into prosody generation system, and thereby the production of more discourse-felicitous prosody. Moreover, given that FDA starts with conceptual planning of message, its integration is particularly suitable for Concept-to-speech systems. As a final note, we suggest that its fundamental rationale is arguably also highly pertinent to Text-to-speech systems, which, however, cannot be elaborated here.

### References

- Baddely, A. (1990) *Human Memory: Theory and Practice*. Lawrence Erlbaum, Hove.
- Chafe, W. (1987) Cognitive constraints on information flow. In R. Tomlin, ed., *Coherence and Grounding in Discourse*. Benjamins, Amsterdam.
- Dirksen, A. (1992) Accenting and deaccenting: A declarative approach. In *Proceedings of COLING 1992*. Nantes, France. IPO Ms. 867.
- Dirksen, A. & Quene, H. (1993) *Prosodic Analysis: the Next Generation*. In "Analysis and Synthesis of Speech", V. van Heuven, & L. C. W. Pols, ed., de Gruyter, Berlin, pp. 131-146.
- Erteschik-Shir, N. (1997) *The Dynamics of Focus Structure*. CUP, Cambridge.
- Garrett, M. F. (1980) *Levels of Processing in Sentence Production*. In "Language Production: Vol. I. Speech and Talk", B. Butterworth, ed., Academic Press, London.
- Gathercole, S. E. & Baddeley, A. D. (1993) *Working Memory and Language*. Lawrence Erlbaum, Hillsdale.
- Grosz, B. & Sidner, C. (1986) *Attention, Intention, and the Structure of Discourse*. Journal of Computational Linguistics, 12, 175-204.
- Hajicova, E. (1987) *Focusing: a Meeting Point of Linguistics and Artificial Intelligence*. In "Artificial Intelligence. Vol. II: Methodology, Systems, Applications", P. Jorrand & V. Sgurev, ed., North-Holland, Amsterdam, 311-321.
- Halliday, M. A. K. (1967) *Intonation and Grammar in British English*. de Gruyter, Berlin.
- Heim, I. (1983) *File Change Semantic and the Familiarity Theory of Definiteness*. In "Meaning, Use and Interpretation of Language", R. Bauerle, Ch. Schwarze & A. von Stechow, ed., de Gruyter, Berlin.
- Ladd, D. R. (1996) *Intonational Phonology*. CUP, Cambridge.
- Levelt, W. J. M. (1989) *Speaking*. MIT Press, Cambridge, MIT.
- Linell, P. (1995) *Troubles with Mutualities*. In "Mutualities in dialogue", Markova, I., C. Graumann & K. Foppa, ed., CUP, Cambridge, pp. 176-216.
- Nuyts, J. (1992) *Aspects of a Cognitive-Pragmatic Theory of Language*. Benjamins, Amsterdam.
- Pierrehumbert, J. (1980) *The Phonology and Phonetics of English Intonation*. Ph.D. dissertation. MIT.
- Prince, E. (1985). *Fancy Syntax and 'Shared Knowledge'*. Journal of Pragmatics, 9, 65-81.
- Reinhart, T. (1981) *Pragmatics and Linguistics: an analysis of Sentence Topics*. Philosophica, 27, 53-94.
- Rochemont, M. (1986) *Focus in Generative Grammar*. Benjamins, Amsterdam.
- Rommetveit, R. (1974) *On Message Structure*. Wiley, New York.
- Sanders, J. & Spooren, W. (1997) *Perspective, Subjectivity and Modality from a Cognitive Linguistic Point of View*. In "Discourse and Perspective in Cognitive Linguistics", W.-A. Liebert, G. Redeker, & L. Waugh, ed., Benjamins, pp. 85-114.
- Sandford, A. J. & Garrod, S. C. (1981) *Understanding Written Language*. John Wiley & Sons, Chichester.
- Taylor, I. & Taylor, N. N. (1990) *Psycholinguistics: Learning and Using Language*. Prentice-Hall International, Inc.
- Tulving, E. (1985) *How Many Memory Systems Are There?* American Psychologist, 40, 385-398.
- Vallduvi, E. (1992). *The Informational Component*. Garland, New York.
- Wilkes, A. L. (1997) *Knowledge in Minds*. Psychology Press, Erlbaum.
- Zuo, Y. (1999). *Focusing on focus*. Ph.D. Dissertation. Peking University, China.