DETERMINING CATEGORICAL STATUS BY FUNCTIONAL CATEGORIES: EVIDENCE FROM MANIPURI AND TAMIL.

Nandaraj Amom
Lecturer/RP,LDC-IL, CIIL, Mysore
nandaraj_amom@yahoo.com
and
Thennarasu Sakkan
Lecturer/RP, LDC-IL, CIIL, Mysore
thennarasu@gmail.com

Linguistic Data Consortium for Indian Languages (LDC-IL)

CONTENTS

- INTRODUCTION
- FUNCTIONAL CATEGORIES (FCS)
- LEXICAL VS FUNCTIONAL
- CATEGORIES AND FEATURES
- CATEGORICAL STATUS
- RESEARCH AREAS AND TEXT ANNOTATION
- CONCLUSION

INTRODUCTION

❑ When lexical elements like nouns and verbs are non-distinct or very similar, particularly in languages like Manipuri and Tamil, the endings make clear what kind of element is used.

❑ Nouns receive endings indicating cases, numbers etc and verb endings indicating aspects, tense, mood etc.

❑ Functional categories only give the lexical ones their categorical status when they are in-between state.

FUNCTIONAL CATEGORIES (FCS)

> multi-dimensional conception of functional categories

> functional categories should be actual words or affixes visible (in our view)

> semantic (gender, number, cases – overt/covert)

> syntactic (agreement in person/number through argument realization)

> pragmatic (discourse markers- topic/focus etc.)

LEXICAL VS FUNCTIONAL

- lexical categories have so called 'descriptive content'

- and the functional categories lack the same

- lexical categories are items constituting the basic units of expression and thought

- functional categories are the ones whose basic role is to mark grammatical/relational features

CATEGORIES AND FEATURES

❖ **ARCHITECTURAL VIEWS:**

❖ **Category in Generative Literature:**

(1) Design Factors

Generative Grammar endorses the hypothesis that language is a component of the human mind/brain: the Faculty of Language (hence ford, FL). Ever since Chomsky (1965: 59), generativist investigation have assumed that FL is largely regulated by three factors:

(1) **FACTORS IN DESIGN**

(a) *Genetic endowment*

(b) *Experience*

(c) *Principles not specific to FL*

[from Chomsky 2005: 6]

(2) STRONGEST MINIMAL THESIS (SMT)

Language is an optimal solution to legibility condition.
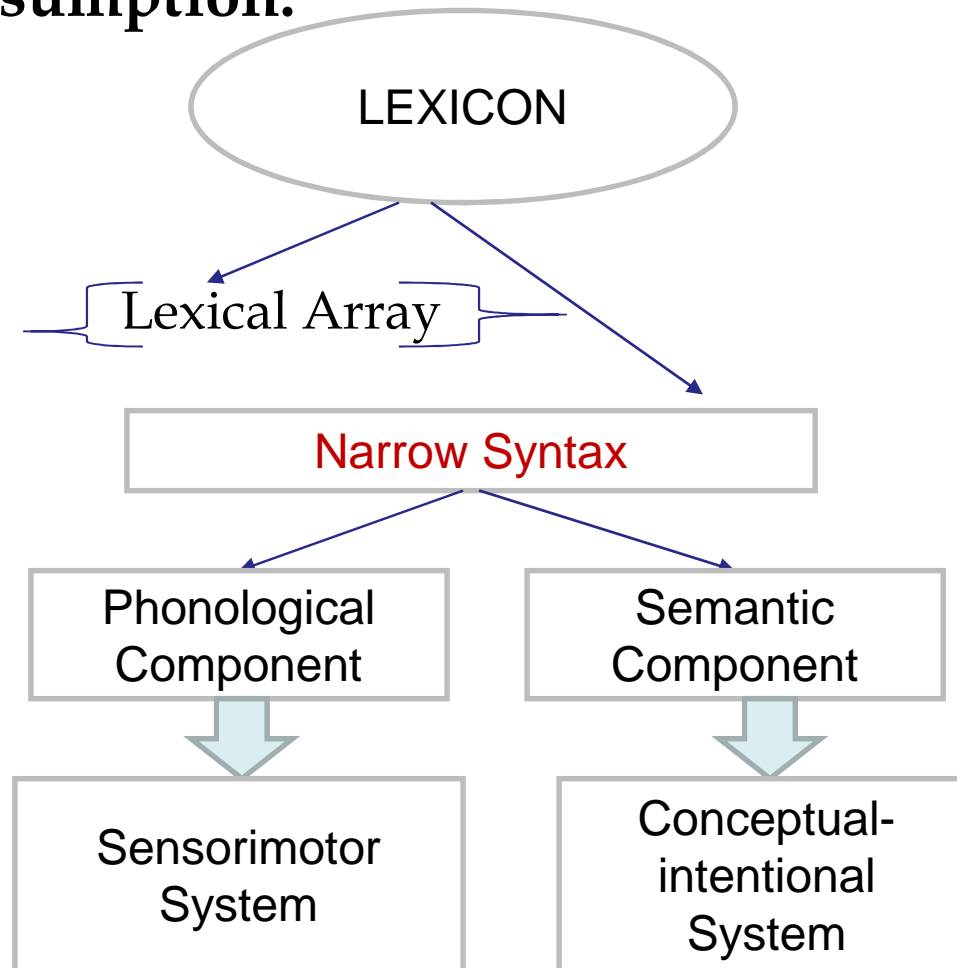
[from Chomsky 2000: 96]

Once (2) is seriously entertained, linguistic inquiry seeks to recast substantive principles from considerations about computational efficiency and properties of the systems with which FL must interact: the sensorimotor (SM) and Conceptual-Intensional (C-I).

More generally, minimalism seeks to show that the basic principles of language [can be] formulated in terms of notions from the domain of (virtual) conceptual necessity (Chomsky 1993a: 171)

- **Minimalist Assumption:**

LEXICON

Lexical Array

Narrow Syntax

Phonological Component

Semantic Component

Sensorimotor System

Conceptual-intentional System

> Now we see that MP regards FL as formed by a Lexicon and a Computational System (i.e., a Narrow Syntax).

  For the time being, let us put Lex to the side (and the important issue of whether it is 'distributed' in Halle & Marantz's 1993 sense), concentrating instead on Narrow syntax.

Computational Operations:

> A language L is taken to consist of two main components: **LEXICON** and **NARROW SYNTAX**

Computational…

The background assumption within minimalism is that FL specifies a set of features available to each particular language L. Let us call this universal repertoire F_{UG} (i.e., features of Universal Grammar). Design factor (1b) makes a selection from F_{FG} to form F_{PG} (i.e., of a Particular Grammar), which are then assembled into a particular lexicon: a collection of Lexical Items (LIs). We then reach the F vs LI distinctions, a non trivial one under the fairly standard ideas that only some features become LIs.

LEXICON ASSUMPTION

➢ **Some features from** F_{UG} become LIs (categorial ones), others are just assembled within LIs, not being manipulated by the based computational operations of Merge.

Hence the lexicon consists of LIs and features, both of which can be accessed without restrictions. However, structure building operations can not treat features as they do LIs (contra Chomsky's 1995b Attract).

Linguistic Data Consortium for
Indian Languages (LDC-IL)

Narrow Syntax sees LIs as atomic units with different properties (i.e., features) assembled within:
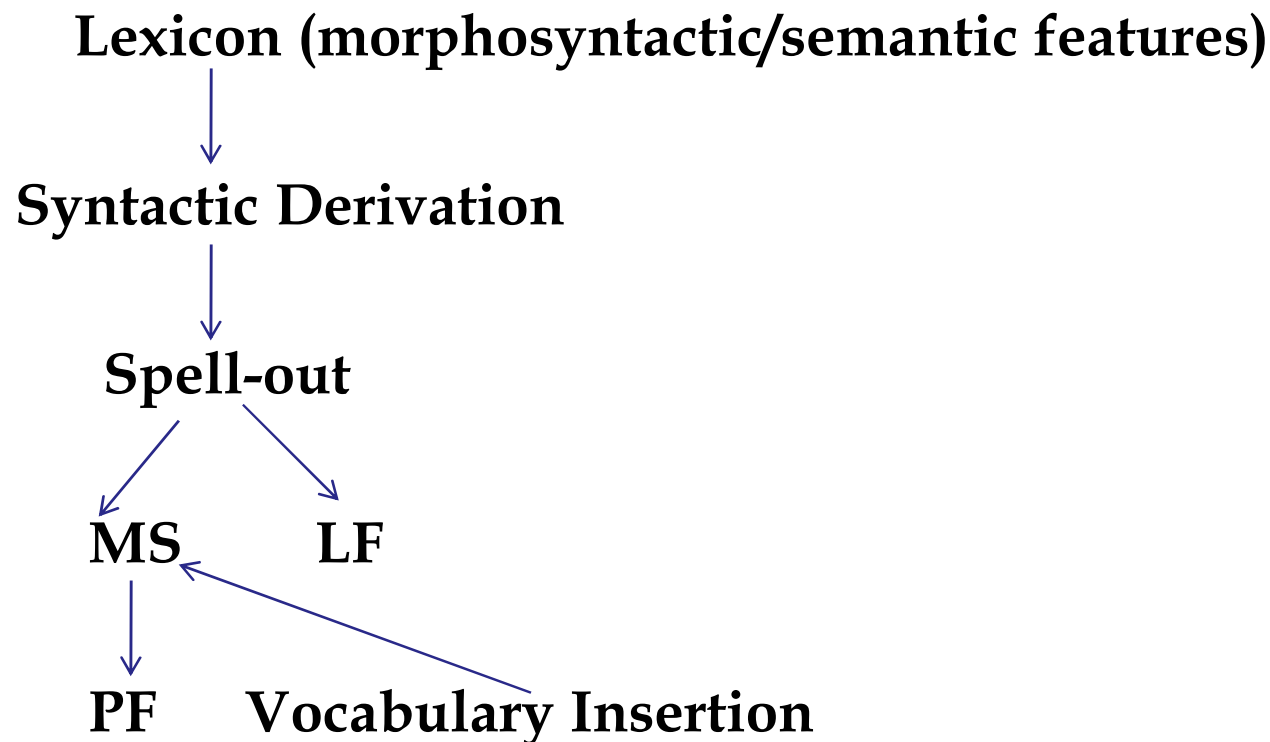
PROPERTIES OF LEXICAL ITEMS (LIs)

Phonological

Semantic

Formal

DISTRIBUTED MORPHOLOGY (DM)

Realizational model of grammar- the morphological component (called Morphological Structure, henceforth MS) operates post-syntactically; syntactic terminal nodes (called morphemes) are supplied with phonological content after syntax (in a process called Vocabulary Insertion):

Lexicon (morphosyntactic/semantic features)

Syntactic Derivation

Spell-out

MS   LF

PF   Vocabulary Insertion

Linguistic Data Consortium for
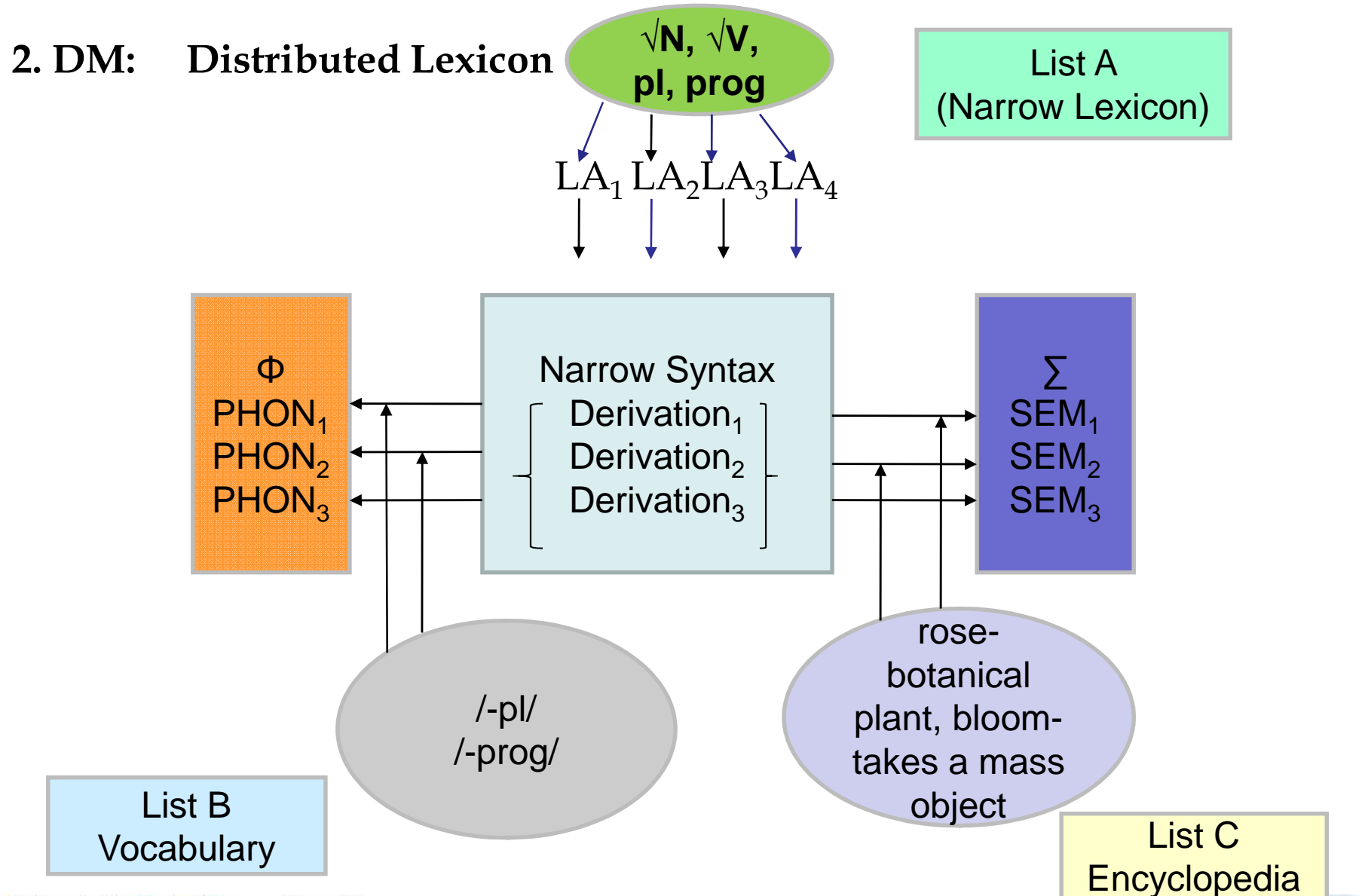Indian Languages (LDC-IL)

Phonological Realization of functional heads

❑ **Substantial lexical heads (N, V, A, P) can be realized by an arbitrary Vocabulary item that matches the category specification (e.g., in Manipuri N may be realized by /ləʲi̥/ , /ciŋ/, /u̥/, /ta etc.; in Tamil N may be realized by /maɾəm/, /kəŋ/, /makan/, etc. dependent on the choice of the speaker).**

❑ **The phonological realization of functional heads is non-arbitrary; it involves a competition between Vocabulary items that are specified for a common subset of the inflectional features in the functional head. The item realizing the greatest number of inflectional features is chosen for insertion.**

❑ **Realization of the Progressive Aspect in Manipuri and Tamil:**

məhak kəp-li (He/she is crying)   vaɾukira ː n (He is coming)

məhak   ca-ri (He/she is eating)   padukkira ː ɭ (She is lying down)

CATEGORIES AND FEATURES

2. DM: Distributed Lexicon

$\sqrt{N}, \sqrt{V},$ pl, prog

List A
(Narrow Lexicon)

$LA_1\ LA_2\ LA_3\ LA_4$

Φ
$PHON_1$
$PHON_2$
$PHON_3$

Narrow Syntax
$Derivation_1$
$Derivation_2$
$Derivation_3$

Σ
SEM_1
SEM_2
SEM_3

/-pl/
/-prog/

rose-botanical plant, bloom- takes a mass object

List B
Vocabulary

List C
Encyclopedia

❖ DM is a framework which abandons traditional conception of the lexicon, exploding the information in (2) in different lists as depicted in (2).

❖ The DM framework suggests that Narrow Syntax makes use of elements from the List A alone: abstract morpho-syntactic features. This makes sense within a system like Chomsky's (1995b, 2000, 2001), where only formal features activate syntax: it is not immediately obvious that phonological properties have any bearing during computation, and things become even trickier when it comes to semantic properties.

❖ Accordingly, increase in paradigmaticity can be systematically linked to the reanalysis of (exponents) lexical items as (exponents of) functional categories.

Distributed Morphology:

➢   No a pre-syntactic lexicon

➢   Roots are category-neutral

➢   No lexical specification of N, A and V as words

➢   the existence of categorizers n for [N], v for [V] etc.

CATEGORIZATION ASSUMPTION: *Roots cannot appear (cannot be pronounced or interpreted) without being categorized: Roots are categorized by merging syntactically with category-defining functional heads. (Embick and Marantz (2006:6).*
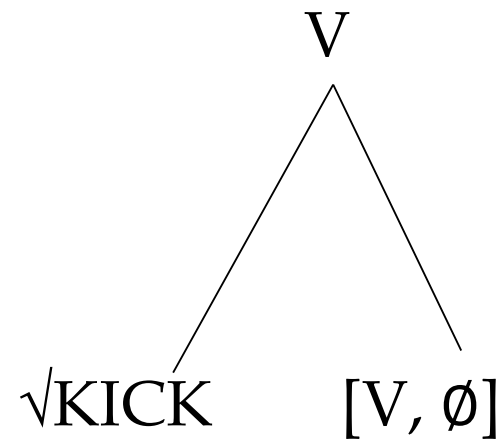
CATEGORIES AND FEATURES

Cat is understood / realized as an N and kick as a V since category neutral √CAT and √KICK merge with the respective category - defining functional heads n and v as in (2)

(1) √CAT as N   (2) √KICK as V
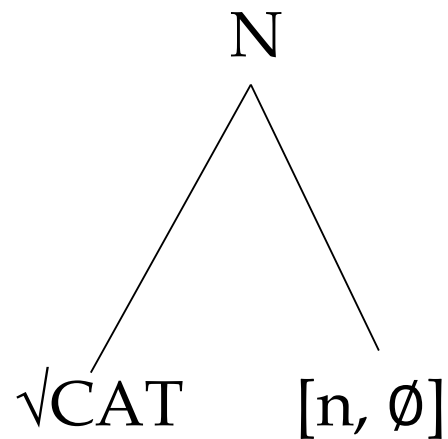
         N   V
        / \   / \
       /   \   /   \
      /   \   /   \
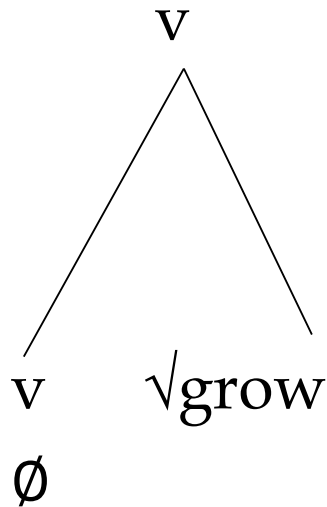   √CAT   [n, ∅]   √KICK   [V, ∅]
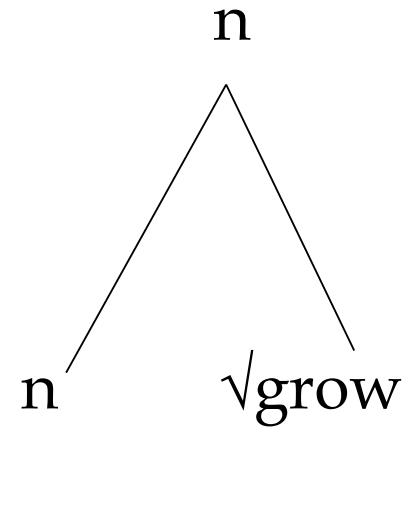
(Embick & Marantz; 2006:5)

CATEGORIES AND FEATURES

❑ <u>**ROOTS in a local relation with the category defining F-MORPHEMES (functional morphemes) v, n and a, respectively**</u>

(3) v grow - Ø   (4)   n grow-th

<pre>
         v   n
        / \   / \
       /   \   /   \
      /   \   /   \
     /   \   /   \
    v   √grow   n   √grow
    Ø   -th
</pre>

√grow as a V in local to v   √grow as a N in local to n

category distinctions

➢ correspond to perspectives on roots or concepts about the world (Baker, 2003:293-294)

➢ A [V] feature imposes an extending-into-time perspective at LF; an [N] feature imposes **a sortal perspective** at LF (Panagiotidis,2005)

➢ we conceive objects and substances as sortal concepts mapping them into nouns extending into space

➢ we conceive dynamic events including activities, achievements and accomplishments as concepts **extending into time** and map them into verbs

✓ categorizes are not functional heads

✓ categorical features impose perspective on the semantic material in the complement of the nP|vP phase

✓ [N] and [V] features behave as phase-edge features [Panagiotidis, 2006]:

A. *phaze-internally, they contribute the interpretive perspective and*

B. *they identify the whole phase externally (as 'nominal' or 'verbal').*

In short, categorizers are the only possible lexical heads

Survive and Features

- The Revised Survive Principle (based on Stroik 1999:286)

- *If Y is a syntactic object (SO) in an XP headed by X, and Y has an unchecked feature [+F] which is incompatible with the feature born on X, Y remains active in the Numeration.*

- To gain an idea of how the computation of human language (C^{HL}) would operate according to the Survive Principle lain out above.

CATEGORIES AND FEATURES

❖ Survive accounts:

❖ the computational system of Human Language requires a Numeration

❖ However, the Numeration must be compiled throughout a syntactic derivation and not selected prior to the derivation.

❖ The Numeration becomes the domain for both the Merge and Remerge operations, thereby obviating the need for the Internal Merge operation.

❖ Furthermore, having a derivationally built Numeration makes it possible to design a crash-proof syntax along the lines developed in Stroik's (2009) Survive-minimalism.

CATEGORIES AND FEATURES

A sample derivational history of a grammatical sentence constructed according to this principle is provided below.

(5) Who cares?
(6) a. Merge {who, cares} → who cares
b. Survive {who}
c. Merge {T, **{who, cares}}** → **T who cares**
d. Remerge {who, {T, **{who, cares}}}** → **who T who cares**
e. Survive {who}
f. Merge {C, **{who, {T, {who, cares}}}}** → **C who T who cares**
g. Remerge {who, {C, **{who, {T, {who, cares}}}}}** → **who C who T who cares**

❑ A CATegory feature, with SUBCATegory features, and Interface features, such as wh-features, case features etc. All these features are SYN features participating in the concatenating operations, Merge and Remerge and which must be checked in the course of a syntactic derivation

CATEGORIES AND FEATURES

OUR VIEWS...

o Roots don't have categorical status.

o There are no visible categorizers, i.e., little x's

o Functional Categories should be interface visible

o The syntactic operations are necessarily correlated with semantic/pragmatic effects

o the syntactic processing advantages of functional categories cannot be easily stated in terms of a single aspect of syntax

o Functional categories only facilitate the overall syntactic processing

o we use only functional categories here, instead of employing morpho-syntactic, morpho-semantic, morpho-pragmatic features

CATEGORICAL STATUS

Manipuri

(5). Root : pa
  pa - siŋ > N
  eyelash PL
  "Eyelashes"

  pa - ri
  read - prog > V
  "/is/am/are reading"

(7). Root: ta
  ta-siŋ-du-nə
  spear-PL-DIST-INST > N
  "with those spears"

  ta-jə-rək-li
  hear-REFL-INCEP-PROG > V
  "has/have been reading"

Tamil

(6). Root: kottu
kottu-kaɭ > N
Bunch-PL
"Bunches"

kottu-kir-atu
peck- PRS – 3SG.NEU. > V
"bird is pecking"

(8). Root: piɻai
  piɻai-kaɭ
  Error-PL > N
  "Errors"

piɻai - tt - aː n
Survive-PST – 3SG. MAS.> V
"He survived"

(9). Root: u

u-siŋ-si-nə

tree-PL-PROX-INST > N

"By these trees"

u-sin-le

see-INward-PRF > V

"have/has been overseen"

(10). Root : u ː ɾ

u:ɾu-kku

village-DAT. > N

"To village"

u:ɾu-kir-atu

Crawl –PRS-3SG.NEU. > V

"Crawl (of certain creatures)"

(11). Root: ka

ka-du-də

room-DST-LOC > N

"In that room"

ka-khət-lu

clim-up-IMP > V

"climp up"

(12). Root : kal

kar-kkaɭ-ai

Stone-PL-LOC. > N

"of stones"

kar-pp-a ː n

Study-FUT.3SG.MAS. > V

"He will Study"

(13). Root: ya
    ya-siŋ-si
    tooth-PL-PRX > N
    "All these teeth"

    ya-re >V
    agree-PRF
"Have/has agree"

(14). Root: **a**ʈi
aʈi-kaɭ-aː l
Foot-PL-INS > N
      "By foots"

aʈi - tt - aː n
Stroke-PST-3SG.MAS > V
"He hit (Stroke with the hand)"

(15). Root : tha
    pa - du
moon-DST > N
"the yonder moon"

tha- ri
plant -prog > V
"is/am/are planting (trees)"

(16). Root: kaː j
kaː j-kaɭ-il
vegitable-PL-LOC. > N
"In vegitables"

kaː j-n̪t-atu
dried-PST-3SG.NEU > V
    "Got dried up"

(17). Root: kha
   kha-rom-də
south-ALLA-LOC > N
"towards the south"

   kha-ri
fence-DCL > V
"is fancing (boundaries)"

(18). Root: iɻai
   iɻai-kaɭ-in-meel-iruntu
Yarn-PL-GEN-PP-PP > N
"on/from the top of Yarns"

   iɻai-tt-aːɾ
Shave-PST-3SG.HUM > V
"(He/She) is shaved"

(19). Root: na
na-si-mak-nə
ear-PL-PROX-EMPH-INST > N
"Even with this ear"

na-gal-li
sick-HAB-DCL > V
   "be sickly"

(20). Root : katai
katai-j-aːl
Story-LOC > N
"By story"

kaɾi-kkir-atu
charred-PRS.PROG.3SG.NEU > V
   "(It) has/been charred"

(21). Root: phəm
    phəm-du-su
    position-DST-INCL > N
       "Also that position/rank"

    phəm-min-nə-ri
       sit-COMI-REC-PROG > V
       "Sitting altogether"

(22). Root : ka ː talɪ
ka ː tali-kk-a ː ka
lover-BEN
"For the sake of lover"

ka ː tali-tt-a ː n
Love-PST-3SG.MAS > V
"He loved"

(23). Root: la
    la-khəi-du
    plantain-CUM-DST > N
"Whatever the plantain leaves concerned"

    la-khət-lək-le > V
    clear-UP-PRF
    "Have/has become clearer"

(24). Root: padɪ
paʈi-kaɭ-ukku
Step-PL-DAT > N
"To the Steps"

paʈi-tt-a ː n-a ː
Read-PST-3SG.MAS.INT
"Has he read?"

Manipuri:

Tomba\NP: n∂ŋ\PPR.2.sg Chaobabu\NP.acc

kh∂ŋbra\NV.nmlz.int ?\PU

Ibomcha\NP: mabudi\PPR.3.sg.def

kh∂ŋŋina\V.dcl.conf

Tomba: Do you know Chaoba?

Ibomcha: As for him, I have already known.

Tamil:

Palani\NP: unakku\PPR. 2.sg.dat Tirumagalai\NP.acc.

teriyumaa\V.int. .\PU

Saritha\NP: avalaippattri\PPR.3sg.mas.acc.PP.ded.

teriyum\V.dcl.

Palani: Do you know Tirumagal?

Saritha: As for her, I have already known.

- In the above dialogue, the discourse marker 'di' 'definitenes' plays a major role in the sense of which the person Tomba is already known to Ibomcha and it imposes a force that it is not necessary for Ibomcha to be asked about the whereabouts of Chaoba.

- This discourse marker is morphologically a simple definite indicating suffix but a linker of discourse world.

- Whether such markers are found within the nominal or verbal is also in the line of investigation by which we can determine the status of the category concerned.

- However, in the case of Tamil, the particle-'pattri' 'dedative' 'about' plays a major role in identifying and recognizing the person concerned. The particle 'pattri' specifies the person referred to in the previous discourse.

SO WE CAN POSIT THE FOLLOWING HYPOTHESIS:

CATEGORICAL DETERMINATION:

DETERMINE THE CATEGORICAL STATUS OF THE ROOT BY THE FUNCTIONAL CATEGORIES IT GOES WITH.

CONCLUSION

- The realizational behavior of Functional Categories presents strong evidence for a gradience view of which it is the functional categories that give the lexical ones their categorical status.

- Different varieties of functional categories can be helpful in the domain of text-annotation such as POS and discourse-act tagging etc.

- It is felt that the automatic machine learning can be mainly applied to the training determination of functional categories in an individual language to produce a huge amount of accurate data which can be further used to induce new patterns.

ACKNOWLEDGEMENTS

We are indebted to Dr. B. Mallikarjun, Head, LDC-IL for his academic moral support and constant encouragement and to our brilliant team that provided us insightful suggestions.

REFERENCES

Abney, Steven Paul, 1987. *The English Noun Phrase in its sentential aspect*. Doctoral Dissertation, MIT.

Baker, Mark 2003. *Lexical categories: verbs, nouns and adjectives*. OUP.

Chomsky, Noam 1970. *Remarks of Nominalization*. In: Jocobs, Roderick, and Rosenbaum, Peter. (eds.) *Readings in English Transformational Grammar*. Waltham, Mass.: Ginn & Company. 184-221.

Chomsky, Noam 1995. *The Minimalist program*. Cambridge, Mass.: MIT Press

Chomsky, Noam 2001. *Derivation by phase*. In Kenstowicz M. (ed.) Ken Hale. *A life in language*. MIT Press. 1-52

Embick, David & Marantz, Alec 2008. *Architecture and blocking*. Linguistic Inquiry 39: 1-53. University of Pennsylvania and MIT.
Embick, David & Noyer, Rolf, 2001. *Movement operations after syntax*. Linguistic Inquiry 555-595

Ferrari, Giacomo, 2004. *State of the art in computational linguistics*. In Linguistics Today – Facing a Greater Cahllenge, eds. by Piet Van Sterkenburg John Benjamins Publishing Company. Asterdum/Philaduphia.

REFERENCES

Fukui, Naoki. 1986. *A theory of category projection and its applications*. Doctoral dissertation, MIT.

Harley, Heidi 2006. *The morphology of nominalizations and the syntax of vP*. To appear in Giannakidou, Anastasia & Monika Rathert (eds.), Quantification, Definiteness, and Nominalization. Oxford: Oxford University Press.

Harley, Heidi 2007. *The bipartite structure of verbs cross-linguistically*, or *Why Mary can't exhibit John her paintings*. Talk presented at the 2007 ABRALIN Congres in Belo

Horizonte, Braizil, March 2007. Heaberli, Eric, 2002. *Features, Categories and the Syntax of A-Positions*: *Cross-Linguistics Variation in the Germanic Languages*. Dordrecht: Kluwer. Jackendoff, Ray 1977. *X'Syntax: A Study of Phrase Structure*. Cambridge, Mass.: MIT Press.

MARANTZ, ALEC, 1997. *No escape from syntax: don't try morphological analysis in the privacy of your own lexicon*. U. Penn Working papers in Linguistics 4: 201-225

MARANTZ, ALEC, 2000. *Words*. Unpublished ms. MIT.

REFERENCES

MARANTZ, ALEC, 2006a. *Seminars delivered at EALING, Paris*. Available form http://web.mit.edu/marantz/Public/EALING/

MARANTZ, ALEC, 2006b. *Phases and words*. Unpublished ms. NYU.

MUYSKEN, PIETER, 2008. *Functional Categories.* Cambridge University Press

PUTNAM, M & STROIK, T, 2008. *Syntactic relation in survive-minimalism*. Ms. Carson-Newman & University of Missouri-Kansas city.

STROIK, T. 1999. *The Survive principle*. Linguistic Analysis 29: 278-303

STROIK T. 2009. *Locality in minimalist syntax*. Cambridge MA: The MIT Press.

THANK YOU

ANY QUERIES!