Transcategoriality: a functional explanation and some typological insights

Stéphane Robert, LLACAN (CNRS & INALCO) robert@vjf.cnrs.fr

1. Introduction

Grammaticalization: - renewed studies, historical perspective (cf. Heine & Kuteva 2002)

Challenging cases: - $t\varepsilon$ in Baka: preposition, auxiliary, co-ordinating or subordinating

conjunction (case marking, subordination, diathesis, predication, derivation, tense-aspect-modality) (Heine & Kilian-Hatz 1994) - *qinnaaw* in Wolof: a noun ("the back"), a preposition ("behind" or

 ginnaaw in violot: a noun ("the back"), a preposition ("benind" or "excepted") or a subordinating conjunction "(causal) since"

(Robert: 1997)

- ba(ng) in Rama "polygrammaticalization" (Craig 1991)

in synchrony \rightarrow "transcategorial functioning"

"transcategorial" morphemes: used synchronically across different syntactic categories ('synchronic grammaticalization')

from grammaticalization to transcategoriality diachronic → synchronic

Crosslinguistically widespread With various degrees and modalities A property of linguistic systems, variously exploited

Questions: - Unity of the morpheme / semantic and syntactic variation?

- Status of the linguistic categories / syntactic flexibility?

- Impact and modalities of transcategorial functioning (typology)

Proposals: - a dynamic model for the analysis of transcategorial functioning

- a typological sketch of transcategoriality

- a few reflexions on the status of linguistic categories

How to account for the semantic and syntactic variation / unity of the morpheme?

2. The proposed model of analysis: "fractal grammar"

2.1. Fractal grammar and fractal functioning (Robert: 1997, 2003a, 2004)

A dynamic model relating the polysemy of these morphemes to their syntactic flexibility:

- (1) the co-text specifies the syntactic category in which the item is used;
- (2) the variation of the syntactic scope of the morpheme produces its polysemy by triggering variation in its semantic scope, and the activation of contextual properties.

Two properties from fractal objects (cf a tree branch, a snow flake...):

SCALE INVARIANCE and SELF-SIMILARITY: a "similar" structure appears at different scales (objects are invariant when undergoing a dilatation) (Mandelbrot 1975)

SCALING LAWS (Sapoval 1997): each scale also has specific scale properties so that there is no strict identity between the same structure appearing at different levels. Rather, we have an 'analogic' structure.

Transcategorial functioning as a fractal property of language:

A similar image-schematic structure functions at different syntactic levels («scales") inside the utterance.

The linguistic 'scale' = the syntactic 'level' at which the unit functions.

The semantics of the morpheme undergoes dilatation of its syntactic scope: scale invariance and scale properties.

Two basic mechanisms:

- construal of a common image-schema («scale invariance")
 - → (semantic) unity of the morpheme
- activation of "scale (or level) properties "
 - → semantic and syntactic variations

The context specifies the level at which this semantic structure functions by defining the syntactic scope of the item.

2.2. Examples

Tupuri (Adamawa, Cameroon) (Ruelland 2003):

kāl "to enter" used as ingressive auxiliary (kàl)

image-schema: intrance in a domain

domain: a place or a process (analogy space / time)

Nêlêmwa (Oceanic, New-Caledonia) (Bril, 2003): roven

roven - verb to finish (1)

- aspectual modifier (terminative) (2)

- nominal quantifier "all" (3)

- (1) co <u>toven</u> o <u>khiiboxa pwaxi-m</u> tavia! 2SG finish LOC beat child-POSS.2SG dog "Stop beating your dog!"
- (2) iwagiik... и keva wany ха keva хe roven 3sg COMP build boat TOP also build finish a... wany hleny boat this.DEICT "He has built up a boat... and this boat is finished"
- (3) hla vhaa agu <u>roven</u> 3PL talk people all "<u>Everybody</u> is talking (about it)"

Image-schema: a totalizing quantification or completed scanning (1) of

(2) the phases of a process (temporal domain) or

(3) the elements of a class

Wolof (Atlantic, Niger-Congo, Senegal): ginnaaw

ginnaaw's senses:

noun - back (body part) (1)

preposition - behind (extended uses: after, except) (2) (3)

subordinating conjunction - since (causal not temporal) (4)

- (1) Jigéén-u Senegaal dañu-y boot seen doom ci ginnaaw woman-CONN Senegal VBFOC.3PL-iMPERF carry their children PREP ginnaaw Senegalese women carry their children on the/their backs
- (2) Mi ngi dëkk ci ginnaaw jàkka ji 3SG...PRESENT live PREP ginnaaw mosque the He lis living behind the mosque
- (4) <u>Ginnaaw</u> Moodu, ñépp ñëw nañu ginnaaw Moodu, all come PRF.3PL Besides (except) Moodu, they all came.
- (3) <u>Ginnaaw</u> faral nga ko, maa ngi dem. ginnaaw to.side.with PRF.2SG him, 1SG...PRESENT go Since you have taken his side, I am leaving

2.3. Scale invariance: the common image-schema

ginnaaw defines an asymmetrical space with a front / back orientation proceeding from a LANDMARK (or LOCATOR) and REFERS TO THE SPACE BEHIND IT (excluding the landmark).

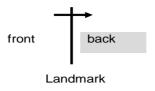


Figure 1: ginnaaw

Which element serves as the landmark (the variable)?

At different levels (scales) inside the utterance:

Landmark = \emptyset the body ginnaaw = noun sense ="the back"

Landmark = a noun ginnaaw = preposition sense = "behind" / "except"

Landmark = a clause ginnaaw = sub. conj. sense = "since"

As a noun (referential scope): ginnaaw refers to the space behind the body.

As a preposition: - ginnaaw refers to the space behind this landmark-noun: SV behind N

- ginnaaw + N: at the front of the clause, the scope of the ginnaaw

phrase is the clause(and not only the predicate):

ginnaaw refers to (thus validates) the 'space' behind the landmark, excluding the landmark; the proposition 'they all came' is true only

behind the landmark 'Moodu'.

Hence the sense 'except Moodu, they all came'.

As a subordinating conjunction: *ginnaaw* expresses a locational relationship between two clauses but not a temporal sequencing (*behind = after P, there is Q).

The clause P ('you have taken his side') is the landmark behind which the clause Q is located, and *ginnaaw* <u>refers</u> to the space behind this landmark. Thus, the main clause ('I am leaving') is the scope of assertion, the focus, and the *ginnaaw*-clause is presented as the starting point of the utterance, a topic (cf constraint on the order):

'Behind (i.e. given) the fact that you have taken his side (P), there is the fact that I'm leaving (Q)'.

ginnaaw validates the main clause as a following consequence of the topic

(sequencing in CAUSALITY and a sequencing IN THE SPEECH ACT; argumentative causality cf Robert 1997)

Image-schema: more precise than semantic bleaching, a semantic matrix form

Tupuri (Ruelland 1998): two nouns > locative preposition "in, inside"

not synonymous

nēn"eye""in, inside"compact domain* in a holebíl"belly""in, inside"hollow interior* in the forest

2.4. Scale properties

Despite a common semantic structure/matrix, the transcategorial morpheme shows different syntactic and semantic properties in its various uses.

The mechanism explaining the (semantic and syntactic) variation between uses:

- the position of the term inside the utterance speficies its categorial status
- its functioning in a specific category triggers the activation of specific properties of this category ("scale properties" i.e. properties of the syntactic level)
- 1. **Triggering factors**: position in the utterance; nature, order of the surrounding terms. e.g. The anteposition of a verb activates its functioning as an auxiliary (with properties) e.g. Position of *ginnaaw* before a noun > preposition; before a clause > subord. conj.
- → Activated properties for each syntactic level:

2. Domain of application and scope of the term

- At the nominal level: referential scope, denotational value; the image-schema is enriched by 2 scale properties of the lexicon: the semantic domain it applies and the "depth dimension" of lexicon (frames, associated scenarii, physico-cultural properties, connotations...)
- These properties are not present in grammatical uses: relational scope; thus the domain on which the image schema is applied is defined by the modified term (cf ginnaaw, roven)

3. Paradigmatic properties

In each use, the term belongs to a different paradigm with specific oppositions contributing to specify its sense (cf *ginnaaw*: paradigm of prepositions *vs.* paradigm of causal subordinating conjunctions: *ginnaaw* (topic) vs *ndax* (focus)

4. Syntactic properties of the structural level

- Nominal phrase: modifiers, complementation, argumental function of the term.
- Verb phrase: aspecto-temporal specifications, nuclear relation to the subjet, valency.... Clause: structure of the predicative relation, saturation...
- Discourse level: a point of view (aspect, perspective), modal values (assertion, interrogation, epistemic status...), a discursive landmark (the topic), a focus; in a complex clause: temporal or causal sequencing, argumentative orientation... cf ginnaaw "since"

5. Semantics of the category, semantics of the function, semantics of the position

- Semantics of the category (noun, adjective, verb): cf. Wierzbicka (1986)
 e.g. noun: "classification", adjective "description"...
- Semantics of the function:
 - e.g. Langacker (1991) subject: 'profiling a primary figure for the predicative phrase'; Croft (1994): 'subject and object': initiator and limit of the causal segment expressed by the verb
- Semantics of the position:
 e.g.French un grand homme (a great man), un homme grand (a tall man);
 Nêlêmwa, quantifiers: anteposition (fraction of discret units) / postposition (globality) cf roven + GN = " all the ", GN + roven = " the whole "

6. Restrictions or loss of combinatory restrictions, specific from the category

e.g. Verb > auxiliary: looses the restriction on the selection of the subject...

7. Scope of anaphora or coreference (specific to the category)

e.g. Converbs grammaticalized in prepositions: loose the constraint of subject coreference: considering his age, he has made excellent progress in his studies. (Haspelmath & König 1995)

NB Scales properties are supposed to be a universal mechanism but their specifications can vary across languages (cf syntactic rules, semantics of positions...)

NB. Theticals (Kaltenböck et al. 2011): the wider scope/scale (+ internal complexity), cf. scope 'beyond the scope of SG' (sentence grammar) 'comments on the utterance or the situation of discourse',

Activation / inhibition / remanence

Hagège (1990: 138)

que tout le monde sorte, les filles <u>exceptées</u> que tout le monde sorte, <u>excepté</u> les filles everyone out, girls excepted

des années <u>durant</u> <u>durant</u> des années during years

2.5. Limits and complexification of the model

Does not account for all cases of polysemy in general: only fractal functioning of grammaticalization: cf metonymy, gramm. of pragmatic inferences of grammaticalization chains

cf. morpheme $t\varepsilon$ Baka

A model to be refined:

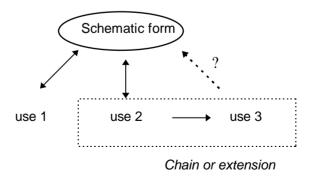


Figure 5: Complexified model

3. Toward a typology of transcategoriality (cf. Robert: 2003b)

A first typological sketch of transcategoriality, based on a collective work: the case studies presented in Robert (2003), and a questionnaire submitted to my colleagues of the LLACAN on:

African languages: Niger-Congo

(Adamawa, Atlantic, Bantu, Gur, Igboïde, Mande, Oubanguian)

Nilo-Saharian (Sara-bongo-baguirmian)

Afroasiatic: Semitic (Maltese), Modern South-Arabic

Oceanic: Nêlêmwa

+occasional incursions in other languages (Basque and Japanese)

Examining: - the relative proportion of transcategorial morphemes

- the nature of the class change (noun to prep., verb to disc.ptcl.)

- the scope of change in each case

- the marking (vs non marking) of class change

- the synchronic *vs* diachronic character of transcategoriality - the morpho-syntactic characteristics of the languages

and morphic symmetric characteristics of the famiguages

First results: different transcategorial functionings and types of languages

- massively vs more restrictively transcategorial languages
- more synchronic vs more diachronic transcategoriality
- oriented (and marked) vs non oriented (direct) transcategoriality
- corresponding to different morpho-syntactic types of languages

"Structural tendencies" to transcategoriality that can be related to the economy of the linguistic systems (different morphosyntactic properties and strategies for the distribution of grammatical information). Various predispositions to transcategoriality in relation to the nature of linguistic system: at least partly predictable.

Three types of transcategorial strategies are distinguished.

3.1. Oriented transcategoriality (= classic grammaticalization)

In languages with heavy morphology (e.g. inflectional languages) (here Hawsa, Maltese, Modern South Arabic languages):

the category change is:

- limited (mainly to the verb)
 (very rare polyfunctionals, if so < other categories: deictics, indefinite pr)
- directed from a source category to a target one (verbs > auxiliaries, sometimes verbes > adverbs or sub.conj., disc.ptcl) (nouns: only body parts > spatial prep.)
- tends to be more diachronic
 (rare polyfunctionals < disappeared nouns; sub.conj < disappeared nouns: cf genitive marking but unclear etymology); freezing

It can be related to a **synthetic and grammatical strategy for the distribution** of syntactic information.

Heavy morphology: inflectional systems, rich system of affixes, gender opposition (M/F), strong verbo-nominal distinction...

The morphology indicates the syntactic category the units belong to

→ category change more difficult but still possible (cf hawsa bayan genitive noun "the back "as sub.conj "after ")

Language units: semantic (notional) units + category indicators + relational nods (preconstructing the syntactic relations with the other elements cf agreement, cases, predicative morphemes...) categories tend to be fixed

→ "oriented" transcategoriality (= classical cases of grammaticalization)
 (limited, from one class to another, more diachronic)

Restricted combinative latitude (categorial rigidity) but synthetic character

3.2. Generic transcategoriality

In languages with light morphology (e.g. isolating languages) (here Banda-Linda, Gbaya, Sängö, Tupuri, Dagara, Ikwere and Nêlêmwa)

- syntactic classes are not or weakly marked (if so, mainly by tonal inflections),
- no conjugations, no nominal inflections, no agreement, no genders
- derivation is limited (vs compounding is highly productive) ...

the category change is:

- massive: numerous transcategorial morphemes * presence of polyfunctionals **
- polydirectionnal (weekly oriented) ***
- unmarked most of the time
- synchronic (and transparent)
- * Examples
- Body parts nouns: used as spatial prepositions but also as morphemes expressing ipseity ("oneself"), reciprocal (Sängö), or temporal or causal conjunctions (Tupuri).
 - Space, movement or time nouns used as demonstratives, locative or benefactive prepositions, valency modifiers, aspectual or modal markers (Banda)...
- ** Frequently: one (or two) "archi-relator" with very variable syntactic scope (introducing complement noun, dependant predicate, relative clause, circumstantial subordinating clause; at utterance level: topic marker (for nouns or subordinating clauses) or focus marker.
 - Cf $n\varepsilon$ Gbaya, nu Ikwere, $m\grave{a}$ modo, \cdot South-Arabic, $s\^{o}$ Sängö, na Dagara, xe and me Nêlêmwa...)
 - Connectors and subordinating morphemes < from other categories (nouns, verbs, adverbs...)

*** Apparently: Any category > Any category

Only for certain languages:

- some blocked paths (Gbaya: * noun > verb / verb > noun)
- privileged paths (verbs > verbal modality, verbs > sub. conj in nêlêmwa)

Sängö (Oubangian, Central Africa): the extreme pole of this typological branch

- units functioning both as: noun and verb; noun and adjective; noun, adjective and adverb; adjective and adverb, noun and quantifiyer; noun and nominal modifier; pronoun, nominal modifier, subordinating morpheme and discourse particle; personal pronoun and predicative morpheme; verb and grammatical marker (person, aspect, modality, valence modifier)...
- one use is hardly "derived" from another
- linguistic units are apparently not categorized or weakly pre-categorized (cf light morphology):
 their syntactic status is specified by the discourse (type-token languages)
- linguistic units: generic notions; can be instanciated in various categories highly combinatorial (compounding rather than derivation)
- cf. lexical compounding (massive):

Gbaya: compouned nouns with $\mathbf{n\acute{u}}$ (Roulon-Doko 2003): egde of a knife, needlepoint, mouth of a basket, side of a field, glowing embers... One use can hardly be derived from another

abstract and generic meaning "the active part of something" referential domain specified by the head noun

Common functional and structural features of these languages:

- (1) no morphological marking of syntactic categories and syntactic relations
- (2) existence of units with a generic meaning underspecified in some aspect (referential domain in the lexicon, syntactic categories in utterance) with a large combinative latitude
 - = semantic matrix whose meaning results from a strong interaction with context
- (3) strong propension to combination (compounding) lexical strategy for the expression of grammatical relations.

Cf Sängö: lexical compounding and lexical strategy for the expression of grammatical rel. cf "syntactic compounding" for verbal modalities: combining full clauses

change of syntactic (and semantic) scope (fractal functioning)

ngû apîka water he-beat "it is raining" zîa ng \hat{u} apîka \rightarrow verb zîa > optative of (ng \hat{u} apîka)

<let-IMPER. water he-beat >

"May it rain"

lo löndö lo nzere

he get-up he please → clause (*lo löndö*) > intensity of (*lo nzere*)

"it is delicious»

ahön ndönî → full clause > adverb

he-overpass the-top

"excessively"

Language units: generic notions: not categorized or only weakly pre-categorized; syntactic status specified by the discourse ("type-token" languages)

Transcategoriality - arises from an initial categorial under-specification

- can be related to an analytical and lexical strategy for the expression of grammatical relations

 → "generic" transcategoriality < initial categorial under-specification (massive, not oriented, synchronic)

Flexiblity of the units, high combinative latitude but more compositionality

3.3. Functional transcategoriality

A third type of transcategorial functioning, examplified by Basque agglutinating language (based on Bottineau 2003)

Basque suffixes: - Traditionally analyzed as "case markers"

(ergative, dative, instrumental, allative, locative-genitive...).

- Express a semantic role (origine, destination..)

- Can apply to various components (various syntactic levels)

e.g. morpheme RA "allative": corresponds to the semantic role "destination"

Infixed in a noun / verb > lexical derivation

ikasle "the one who learns" = the student

iRAkasle "the one who (gives to) learn >= "the teacher"

ikusi " to see "

eRAkutsi " to (give to) see " = " to show "

Suffixed to an argument > directional complement

nire etxeRA noa <my house-to go > "I go back home"

Suffixed to a predicate > destination

eRAkutsi " to show "

[eRAkuste]RA noa "I'm going to show (you) "

Suffixed to a demonstrative > distal demonstrative

hau, hori, hura < hau-RA

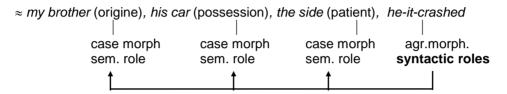
e.g. morpheme k: "origine"

- scope on argument: causal origine of a process (agent)
- spatial or temporal origine of a process
- locative ko
- scope on a clause: conditionnal (*ko* unaccomplished conditonnal), *ke* the conclusive conditionnal): the clause with *k*: origine of the following clause

This transcategorial functioning is permitted by the system of Basque language, which combines two distinct processes:

- (1) case markers: semantic roles, "semantic topicalization" of the components
- (2) agreement morphemes on the predicate: specifiying their syntactic rôle:

" My brother crashed the side of his car "



The **dissociation** between semantic and syntactic roles allows the "case markers" to function with different components, at different syntactic levels.

Transcategoriality < not from category change / crossing

(as for oriented transcategoriality)

< not from category specification in discourse

(as for generic transcategoriality)

but < functional distribution of semantic and syntactic roles

It corresponds to a **selective and distributed strategy for grammatical information** (semantic roles and syntactic roles are expressed by distinct units). Due to this functional distribution, the morphemes expressing semantic roles can apply to various syntactic structures.

This kind of transcategoriality can be called "functional".

Conclusion: importance of the degree of morphosyntactic freedom of the grammatical markers in the propension to transcategoriality; distribution of the grammatical information: dissociation between conceptual components and relational components into different units.

The morphosyntactic status of the grammatical "units" can vary (lexical units in grammatical use as in Sängö, affixes freed from syntactic roles as in Basque, tones as in Tupuri or Gbaya) but the point is that they tend to be autonomous units.

The more autonomous are the grammatical markers (analytical strategy), the easier are the category changes for linguistic components.

4. On the status of linguistic categories

linguistic categories: fuzzy categories (prototype model) or

emergent categories (Hopper 1987, Bybee & Hopper 2001)?

depends on the level of analysis we are considering:
the pattern of the language system, or the way the categories work in

discourse.

Transcatégorial functioning: categorial flexibility (constructed in discourse)

the category pre-exists in the system as a model of functioning.

References

Bottineau, Didier, 2003, Syntaxe génétique et typologie cognitive: la genèse des énoncés basque, anglais et japonais, 10ème Colloque International de Psychomécanique du Langage, Oloron-Sainte-Marie.

Bril, Isabelle, 2003, Quantification, aspect et modalité: phénomènes de portée et d'échelle, quelques exemples en nêlêmwa, in S. Robert (éd): 53-68.

Broschart Jürgen, 1997, Why Tongan does it differently: Categorial distinctions in a language without nouns and verbs, *Linguistic typology* 1, 123-165.

Croft, William, 1994, The semantics of subjecthood. In Marina Yaguello ed., Subjecthood and Subjectivity, Paris, Ophrys, 29-76.

Craig Colette G., 1991, Ways to Rama: A case study in polygrammaticalization. In Elisabeth C. Traugott and Bernd Heine eds, *Approaches to grammaticalization*, Volume 2. Typological studies in language 19. Amsterdam and Philadelphia, John Benjamins: 455-492.

Diki-Kidiri, Marcel, 1995. Le sango, in R. Boyd (ed.) *Le système verbal des langues oubanguiennes*. Coll. Lincom Studies in African Linquistics, 07 Editions LINCOM Europa: 1141-164.

Diki-Kidiri, Marcel, 2003. Le déictique sô du sango, in S. Robert (éd): 189-202.

Hagège, Claude, 1990, *The Dialogic Species. A Linguistic Contribution to the Social Sciences.* New-York. Columbia University Press.

Hagège, Claude, 1993. *The Language Builder*. Current Issues in Linguistic Theory 94. Amsterdam and Philadelphia: John Benjamins.

Heine Bernd & Christa Kilian-Hatz, 1994, Polysemy in African languages: An example from Baka, Thomas Geider & Raimund Kastenholz (eds), Sprachen un Sprachzeugnisse in Afrika. Eine Sammlung philologischer Beiträgen Wilhelm J.G. Möhlig zum 60 Geburtstag zugeeignet, Köln, Rudiger Köppe Verlag.

Heine, Bernd & Tania Kuteva, 2002. World Lexicon of GRammaticalization. Cambridge: C.U.P.

Hopper P. 1987, Emergent Grammar, BLS 13: 139-157.

Kaltenböck, Gunther, Bernd Heine & Tania Kuteva. 2011. On thetical grammar. *Studies in Language* 35, 4: 848–893.

Langacker Ronald W., 1991, Foundations of Cognitive Grammar, Vol. 2., Standford, California: Standford University Press.

Mandelbrot, Benoît, 1975. Les Objets fractals, forme, hasard et dimension. Paris: Flammarion.

Robert, Stéphane, 1997, "From body to argumentation: grammaticalization as a fractal property of language (the case of Wolof ginnaaw)", *Berkeley Linguistics Society, BLS* 23S, 116-127.

Robert, Stéphane (éd.), 2003. Perspectives synchroniques sur la grammaticalisation: Polysémie, transcatégorialité et échelles syntaxiques, Collection Afrique et Langage, Editions Peeters: Louvain.

Robert, Stéphane, 2003a. De la grammaticalisation à la transcatégorialité: Polygrammaticalisation, grammaire fractale et propriétés d'échelles, in S. Robert (éd.) (sous presse).

Robert, Stéphane, 2003b. Vers typologie de la transcatégorialité, in S. Robert (éd.), *Perspectives synchroniquessur la grammaticalisation: Polysémie, transcatégorialité et échelles syntaxiques*, Collection *Afrique et Langage*, Editions Peeters: Louvain.

Robert, Stéphane, 2004, The challenge of polygrammaticalization for linguistic theory: fractal grammar and transcategorial functioning", in Zygmunt Frajzyngier, Adam Hodges and David S. Rood eds., *Linguistic Diversity and Language Theories* [Studies in language companion series 72],J. Benjamins: Amsterdam / Philadelphia, 119-142.

Roulon-Doko, Paulette, 2003. Les parties du corps et l'expression de l'espace, in S. Robert (éd.): 69-84.

Ruelland, Suzanne, 1998, Je pense et je parle comme je suis: le corps, le monde et la parole en tupuri, *Faits de langues* 11-12, 335-358.

Ruelland, Suzanne, 2003. Verbes, auxiliaires et déplacements dans l'espace en tupuri, in S. Robert (éd.): 127-148.

Sapoval, Bernar, 1997. Universalité et fractales. Paris: Flammarion.

Wierzbicka, Anna, 1986, What's a noun? (or: how do nouns differ in meaning from adjectives?), Studies in Language 10: 353-89.