

Variable Negation of Adjectival Predicates in Moroccan Arabic: A Large-Scale Corpus Study

Ali Nirheche & Abdelkader El-Khaoua

Background: negation in Moroccan Arabic

Moroccan Arabic has two negation strategies, **mafi** and **ma-...-f**, sensitive to predicate type:

- Circumfix **ma-V-f** for verbal predicates (Benmamoun 2000, Ouhalla 1997)
- Particle **mafi** for non-verbal predicates (Harrell 1962)

- (1) a. Ali ma-ktəb-f
Ali Neg-wrote.3ms-Neg
'Ali didn't write.'
- b. Ali mafi muʕəllim
Ali Neg teacher
'Ali is not a teacher.'

The puzzle: one category, two negators

Adjectives complicate the picture:

- Both negators attested with adjectival predicates (Benmamoun et al. 2013, Lafkioui 2013, Soufi 2020)
 - (2) a. Ali ma-zwin-f
Ali Neg-handsome-Neg
'Ali is not handsome.'
 - b. Ali mafi zwin
Ali Neg handsome
'Ali is not handsome.'
- Lafkioui (2013): variation is contextual — ma-...-f for disagreement, mafi for neutral

Research question & hypothesis

Research question

- Is the choice between **maji** and **ma-...-j** with adjectives free, contextual, or structural?

Adjective type as a factor governing this variation:

- Non-verbs evolve to behave like verbs (Cohen 1975, 1984; Clark & Clark 1979)
- A subset of adjectives, adjectival participles, are derived from verbs: they contain a verbal head (Kratzer, 1994; Embick, 2004)
 - We expect **adjectival participles** to prefer **ma-...-j**
 - **Plain adjectives** should retain **maji**

Data source: the Atlaset corpus

Atlaset (Bounhar & Majjodi, 2025)

- Arguably the most comprehensive publicly available collection of Moroccan Arabic text
- **Sources:** news portals, local forums, cultural blogs, personal narratives, social-media posts, user-generated content — all in Arabic script.
- ~1.17M rows, ~155.5M tokens

Corpus study: Identifying negation

Python script to find and categorize all sentences with the two negators:

- **ma-...-j**: word starts with [ma-] ما and ends with [-j] ش
- **maji**: exact-match search for the word maji ماضي.
 - **Result:** 1,182,190 negated sentences

To extract negated adjectives automatically, we built a Moroccan Arabic adjective lexicon:

- Manually compiled all adjectives in Harrell & Sobelman's (1966) Moroccan Arabic–English dictionary → 839 adjectives
- The set was converted from Latin → Arabic script (Python), then manually checked and corrected.
- Expanded coverage with morphophonological variants not in the dictionary:
 - Feminine forms (dictionary entries are masculine only) → suffix -a (/ ة)
 - Orthographic variants, e.g. /ð/ (ذ) often written as /d/ (د)

Corpus study: Extracting negated adjectives & final dataset

Matched lexicon entries against sentences with a neighboring negator (mafi or ma-f):

- **35,066 tokens** of negated adjectives (~3% of all negated sentences)
- **406 unique adjectives** (48% of the lexicon)
- Near-even split overall: 18,549 ma-...-f (53%) | 16,517 mafi (47%)

Adjectival participles in our data

Treated as adjectives: language lacks non-finite verb forms (Ennaji 2004)

- Agree with subject in gender and number, not in person
- Occur in attributive and predicative positions, like other adjectives
- **Classification** = adjectives matching the Harrell (1962) patterns below; all others treated as **plain adjectives**

Active Pattern	Tokens	Passive Pattern	Tokens
<i>CaCəC</i>	12,035	<i>mCaCi</i>	962
<i>CaCi</i>	982	<i>məCCuC</i>	920
<i>CaC</i>	423	<i>mCaCəC</i>	629
		<i>məCCəCCəC</i>	318
		<i>məCCaCəC</i>	25
		<i>mCəCCəC</i>	22
		<i>məCCaC</i>	17
		<i>muCaC</i>	5
		<i>maCi</i>	5

Overall distribution

35,066 negated-adjective tokens — 406 unique adjectives. Both negators are common; 60% of adjectives are restricted to one negator.

Negation strategy	Count	Example
Only with ma-...-f	36	zərban ‘rushed’
Only with mafi	204	tʰəmmaʃ ‘greedy’
Variable (both)	166	məzjan ‘good’

Adjective	ma-...-f %	mafi %
mətʰəkkəd ‘sure’	53%	47%
msali ‘available’	99%	1%
darori ‘necessary’	1%	99%

Main result: adjective type predicts the choice

The two adjective classes show almost the opposite pattern:

Adjective type	ma-...-j	mafi	Total
Adjectival participles	13,787 (84.4%)	2,556 (15.6%)	16,343
Plain adjectives	4,762 (25.4%)	13,961 (74.6%)	18,723

Poisson mixed-effects models

Two Poisson mixed-effects models in *lme4* — adjective type as fixed effect (plain = baseline); random intercept per unique adjective.

Model 1 — predicting mafi counts

Predictor	Estimate	SE	z	p
(Intercept)	1.906	0.104	18.30	<.001 ***
adj_participle	-0.43	0.211	-2.03	.042 *

Model 2 — predicting ma-...-f counts

Predictor	Estimate	SE	z	p
(Intercept)	-1.328	0.28	-4.74	<.001 ***
adj_participle	+2.545	0.411	6.19	<.001 ***

Being a participle significantly raises ma-...-f and lowers mafi.

Lafkioui (2013): is variation contextual?

- Proposal: **ma-...-f** marks **disagreement**; **mafi** marks **neutral** context.
- Coded 5,512 disagreement tokens (sentences with denial particle *lla* 'no'); restricted to 166 variable adjectives.
- **ma-...-f** accounts for 71.8% of tokens in disagreement contexts and 53.2% in neutral contexts.
- This gap is an **artifact**: participles overrepresented in disagreement contexts (67.4% vs 46.2%).

Adjective type	Disagr. ma-...-f	Disagr. Total	Neutral ma-...-f	Neutral Total
Adjectival participles	3,314 (93.9%)	3,528	10,378 (82.3%)	12,605
Plain adjectives	441 (25.9%)	1,705	4,144 (28.2%)	14,670

Control for adjective type → context effect disappears for plain adjectives.

The asymmetry to explain

Strong preference on opposite ways

- Adjectival participles — **ma-...-f** in **84.4%** of tokens
- Plain adjectives — **mafi** in **74.6%** of tokens

Neither preference is categorical

- **15.6%** of adjectival participles take **mafi**
- **25.4%** of plain adjectives take **ma-...-f**

Lexical variation among participles

- *msali* ‘available’: 99% **ma-...-f**; *mɕqul* ‘reasonable’: 1% **ma-...-f**

A diachronic change in progress

Alluhaybi's (2019) Arabic Negative Cycle: dialects evolve through stages

- **Initial stage:** one negator negates both verbal and non-verbal clauses
- **Intermediate stage:** verbal/non-verbal split (ma-...-f vs. mafi). **MA classified here.**
- **Later stage:** non-verbal negator (e.g., Cairene **miš**) expands **into verbal** domain

But MA diverges from the predicted trajectory: ma-...-f extending its domain

- **Lafkioui (2013):** ma-...-f incorporating adjectival, locative, existential predicates
- Similar observations made by recent work (e.g., Chatar-Moumni, 2013; Benmamoun et al., 2013)

Negation variation and change in Amazigh (Bensoukas 2012)

- Synchronic dialectal variation in Amazigh negation is also a case of diachronic change in progress.
- Some Tashlhiyt varieties lose negative verb morphology; Tarifit & Figuiq extend it
- Difference from our case: Amazigh change concerns the *form of the negator* (verb-stem morphology, particle system); ours concerns *which predicative heads* negation attaches to

Innovation in the negator, not the predicate

Alluhaybi (2019)

- Non-verbal items (e.g., preposition *çind* ‘at / have’) acquire verbal syntax → switch to **verbal** negation

Adjectival participles look like the same case, but they already contain a verbal core (v; Kratzer 1994, Embick 2004, Doron 2006), so no reanalysis of the participle is needed.

- The innovation lies in the negator: *ma-...-f* is loosening from [+V] to any predicative head [Pred]:
 - Adjectival participles were already [+V] (via their v head), so they were eligible for *ma-...-f* before the loosening.
 - The loosening is reaching predicates with no verbal structure: plain adjectives, which only [Pred] can license
 - So participles lead because they already qualified; plain adjectives lag because they depend entirely on the innovation

Multiple Grammars framework

The main idea (Kroch 1989; Yang 2003; Roeper 1999; Amaral 2014)

- Learners possess a **finite inventory** of competing grammars $G = \{G_1, \dots, G_n\}$
- Each grammar carries a probability weight $p_i(t)$, with $\sum p_i = 1$ and $0 \leq p_i \leq 1$

How weights move through learning

- On each input $s \in E$, weights are **rewarded** if the grammar parses s , penalized if it fails
- Penalty $c_i = Pr(G_i \text{ fails to parse } s \mid s \in E)$: lower penalty $\Rightarrow$ grammar reinforced

Linguistic change as weight redistribution

- Diachronic change = **shift in distribution** across grammars over generations of learners
- Variation in the input shapes the next generation's weights; ongoing change is built in

Two competing grammars for ma-...-ʃ

- **Grammar A (Conservative):** ma-...-ʃ requires a **[+V]** head; licenses verbs and adjectival participles whose v is still accessible
- **Grammar B (Innovative):** ma-...-ʃ requires a **[Pred]** head; licenses any predicative element, verbal or not

How MG explains the asymmetry

Grammar	Constraint on ma-...-f	Adjectival participle	Plain adjective
A (Conservative)	Requires [+V]	Variable (ma-f if v transparent; mafɪ if lexicalized)	Blocked
B (Innovative)	Requires [Pred]	Accepted	Accepted
Corpus result	—	High ma-...-f (84.4%)	Low ma-...-f (25.4%)

- The 15.6% of adjectival participles taking **mafɪ** are **lexicalized** participles where **v** is opaque: only Grammar B could license them, and Grammar B's weight is still low
- Why loosening happens for plain adjectives: **analogical extension** (Kodner 2022); learners observe participles taking **ma-...-f** and generalize to the broader adjective category

Lexical diffusion across participles

- Per-lexeme verbal-negation rates span the full 1%–99% range, i.e., **lexical diffusion**
- Items where **v** stays transparent pattern with verbs; items where the participle has **lexicalized** behave like plain adjectives

Class	Participle	ma-...-f	mafi	% verbal	Gloss
Transparent	<i>msali</i>	951	11	99%	'available'
Transparent	<i>qari</i>	226	5	98%	'educated'
Variable	<i>waḍəḥ</i>	345	290	54%	'clear'
Variable	<i>mətʔəkkəd</i>	153	138	53%	'sure'
Lexicalized	<i>mʕqul</i>	7	530	1%	'reasonable'
Lexicalized	<i>məʃhur</i>	1	39	2%	'famous'

The change isn't happening at the category level: it's happening lexeme by lexeme.

Conclusion

Finding: Adjective negation in MA is not free variation: it is conditioned by **adjective type**.

Theoretical contribution

- The innovation lies on the **negator**, not the predicate: *ma-...-f* relaxes its selection from [+V] to predicative heads.
- Multiple Grammars accounts for the gradient: lexical diffusion, lexicalization, and grammar competition.

Implications

- Refines Alluhaybi's (2019) cycle: MA diverges from the predicted trajectory.
- Demonstrates the power of large-scale corpus methods for capturing patterns of variation and change.

Limitations

- The diachronic claim is based on convergent and cross-dialectal evidence, not direct apparent-time or real-time data.
- Item-level variation among participles is real, but *why* some lead and others lag is open — morphological transparency is one proposal; frequency or semantic factors may also play a role.
- Analysis is limited to adjectives; extending *ma-...-f* to pronominals and nouns remains open.

Thank You