Surveying Perspectives on Language Technology and the Perceived Needs of Moroccan Arabic and Tamazight Speakers

Avery C. Kanel[1,2,*] **Christian Schuler**[1,2,*] **Yusser Al Ghussin**[1,2] **Yassine Chaouri**[3]
Bader Ait Ammi[4] **Simon Ostermann**[1,2] **Josef van Genabith**[1,2]

[1]Saarland University, [2]German Research Center for Artificial Intelligence (DFKI),
[3]National Institute of Posts and Telecommunications, Rabat,
[4]Université Sultan Moulay Slimane, Beni Mellal, [*]Equal contribution

30th Conference of the Foundation for Endangered Languages
Paris, 3–5 November 2026

The Team



Avery C. Kanel



Christian Schuler



Yusser Al Ghussin



Yassine Chaouri



Bader Ait Ammi



Simon Ostermann



Prof. Dr. Josef van Genabith

Question of Interest

What do speakers of Moroccan Arabic (Darija) and Tamazight want from language technology?

We received responses from **29 speakers** how they use their languages, how they view variation and standardization, and which technologies they would find useful.

A Layered Language Ecology

1 Diglossia

Modern Standard Arabic (MSA) and French hold the written and prestigious domains.
(Fathi, 2024; Hermas, 2025)

2 Official Languages

The official languages of Morocco are MSA and Tamazight.
(Falchetta and Guerrero, 2023; Heath, 2013)



Map: *Languages of Morocco* (CC0)

The problem: Outside Europe, very few NLP community needs surveys exist for unofficial and minority languages, and none has engaged speakers across Morocco.

Survey Design

1 Localize

Machine translated into French and MSA. Native speakers translate into Darija and Tamazight.

2 Screen

Keep only speakers of Darija or a Tamazight variety.



Note on Inclusivity: Respondents named their own language and where they speak it. Also, we do not presuppose desirability of digitization or standardization.

Ethics & Data Collection

Anonymous

Responses were collected anonymously; all but the first two questions could be skipped.

Voluntary

Participation was voluntary and not tied to a reward of any kind. (Millour, 2019)

Reciprocal

Findings go back to the communities through the channels used to advertise the survey.

Target Questions

Use

Which languages they use, how often, and whether they write them.

Q4: *"Which other languages do you use every day? For example, at school, at work, or with your family."*

Beliefs

Views on variation, writing, and standardization.

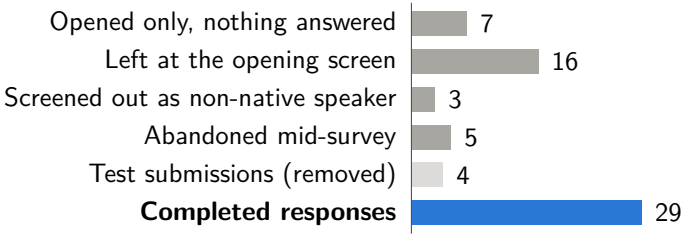
Q14: *"My language should have a standard spelling."* (agree / disagree)

Tech

Which technologies should exist, and what they already know of.

Q18: *"There should be spell- and grammar checkers for my language."*

Who Took Part?



64 recorded survey sessions as of 14 September 2026

How they found us

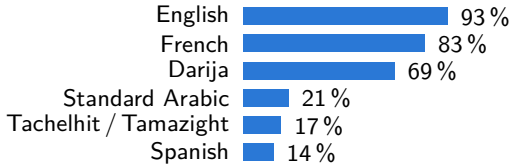
Social media	12
Acquaintances	7
Researchers	5

Median time: **9.2 min**

Informal channels reach speakers better (Millour, 2019).

Who Are the Respondents?

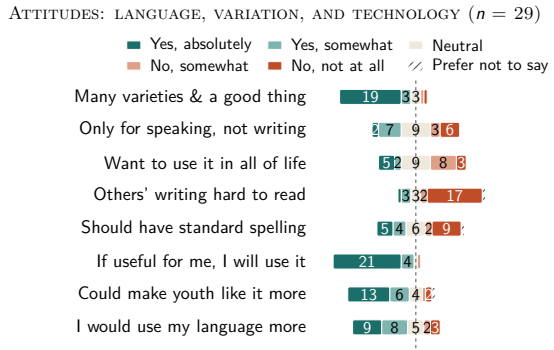
Other daily languages (multiple answers)



	n	%
Age		
≤19	9	31
20–29	16	55
30–39	2	7
Region		
Casablanca	8	28
Oriental region	2	7
other (1 each)	9	31
unspecified / no answer	10	34

Note on Sub-population: 93 % use English daily vs. a national 20.5 % of the literate population (Haut-Commissariat au Plan, 2024): young, educated and online.

Attitudes Toward Language



At a glance

93 % would use a technology they found useful.

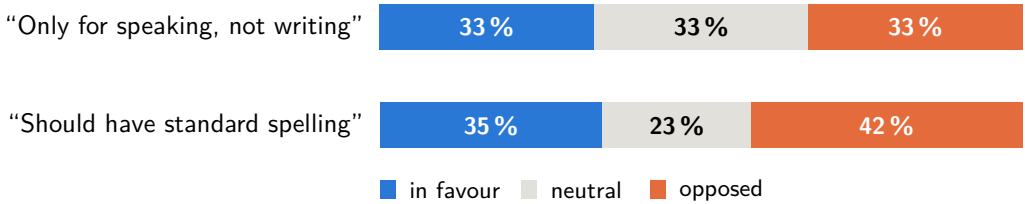
81 % see the varieties of Darija as a good thing.

73 % think more technology could make it more appealing to youth.

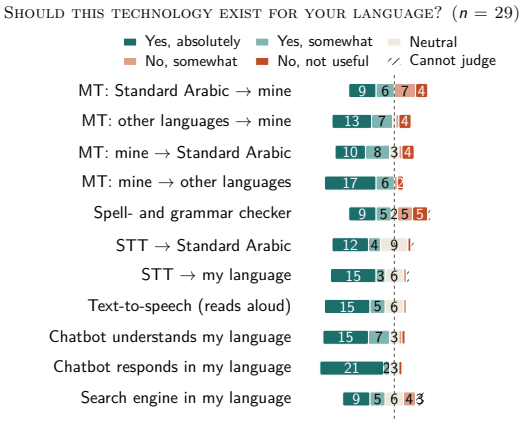
63 % would use their language more.

Q 14, Q 26; of those answering

Standardization Divides Respondents



Which Technologies Should Exist?



At a glance

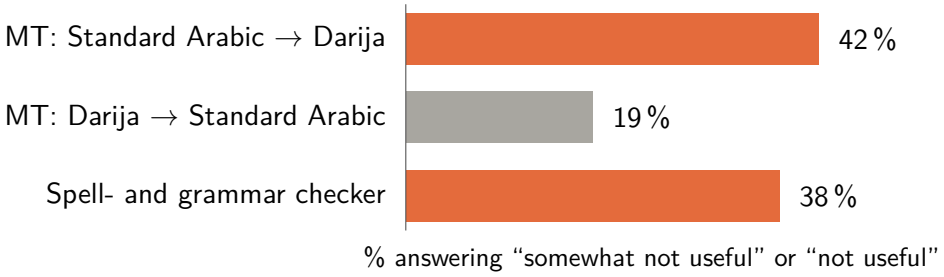
Chatbots in Darija are the most wanted: over 80 % in favour.

Speech tech draws almost no negative answers: STT 62–72 %, TTS 74 % positive.

Least wanted: MT from Standard Arabic into Darija, and spell checkers.

Q 16–Q 23

Less Desirable Technologies



Spelling in Darija

MT

Respondents lean slightly towards not caring which spelling is used: **10** don't mind vs. **8** who want their own.

Chatbots

Respondents lean towards wanting their own spelling: **14 of 27** want this. Only **4** don't mind.

Script

79 % write their language: **66 %** in Latin script, **14 %** in Arabic script.

Q 9 and Q 29

Low Awareness: 78 % know of no Darija language technology.
Those that do name genAI: *"kandn chatGPT kay3rf yktb b darija"*
("I think ChatGPT knows how to write in Darija").

Discussion

Popular Tech

Most view language technology positively. Darija chatbots are the most popular.
(Shang et al., 2025)

Unpopular Tech

MT from Standard Arabic into Darija and spell checkers are the clear exceptions.

Speech

Speech-to-text and text-to-speech are about equally popular, unlike earlier dialect studies.
(Blaschke et al., 2024)

Limitations & Outlook

Limitations

- The largest group of located respondents lives in Casablanca-Settat (3.6 % mother-tongue Tamazight vs. 18.9 % nationally).
- Online recruitment favours the region with the highest internet use.

(Haut-Commissariat au Plan, 2024; Caubet, 2017)

Outlook

- Reach Tamazight-first speakers.
- Target regions with a large Amazigh population.
- Transform the survey into a non-digital format.
- Report findings back to the communities through the same channels.

References I

Steven Bird. Local languages, third spaces, and other high-resource scenarios. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 7817–7829, Dublin, Ireland, 2022. Association for Computational Linguistics.

Verena Blaschke, Christoph Purschke, Hinrich Schütze, and Barbara Plank. What do dialect speakers want? A survey of attitudes towards language technology for German dialects. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 823–841, Bangkok, Thailand, 2024. Association for Computational Linguistics.

Dominique Caubet. Darija and the construction of "Moroccanness". In Reem Bassiouney, editor, *Identity and Dialect Performance: A Study of Communities and Dialects*, chapter 6. Routledge, London, 2017.

Grzegorz Chrupała. Putting natural in natural language processing. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 7820–7827, Toronto, Canada, 2023. Association for Computational Linguistics.

Ahmed Ech-Charfi and Youssef Bouaddi. The receptive vocabulary size of Moroccan learners of Modern Standard Arabic. *Journal of Applied Language and Culture Studies*, 6(3):1–19, 2023.

Jacopo Falchetta and Jairo Guerrero. Koineization trends in spoken Arabic across three Moroccan towns. *Kervan – International Journal of Afro-Asiatic Studies*, 27(1), 2023.

Said Fathi. Dual diglossia in Morocco: A new sociolinguistic situation. *International Journal of Language, Linguistics, Literature and Culture*, 3(4):1–9, 2024.

Atiqa Hachimi. The urban and the urbane: Identities, language ideologies, and Arabic dialects in Morocco. *Language in Society*, 41(3):321–341, 2012.

Haut-Commissariat au Plan. Recensement général de la population et de l'habitat 2024, 2024.

Jeffrey Heath. *Jewish and Muslim Dialects of Moroccan Arabic*. Routledge, London, 2013.

Abdelkader Hermas. Managing Morocco's linguistic market: A linguistic citizenship proposal. *European Journal of Language Policy*, 17(2):237–254, 2025.

References II

Zoey Liu, Crystal Richardson, Richard Hatcher, and Emily Prud'hommeaux. Not always about you: Prioritizing community needs when developing endangered language technology. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 3933–3944, Dublin, Ireland, 2022. Association for Computational Linguistics.

Alice Millour. Getting to know the speakers: A survey of a non-standardized language digital use. In *9th Language & Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics*, Poznań, Poland, 2019.

Guokan Shang, Hadi Abdine, Yousef Khoubrane, Amr Mohamed, Yassine Abbahaddou, Sofiane Ennadir, Imane Momayiz, Xuguang Ren, Eric Moulines, Preslav Nakov, Michalis Vazirgiannis, and Eric Xing. Atlas-Chat: Adapting large language models for low-resource Moroccan Arabic dialect. In *Proceedings of the First Workshop on Language Models for Low-Resource Languages*, pages 9–30, Abu Dhabi, United Arab Emirates, 2025. Association for Computational Linguistics.

Claudia Soria, Valeria Quochi, and Irene Russo. The DLDP survey on digital use and usability of EU regional and minority languages. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan, 2018. European Language Resources Association (ELRA).

Full Profile of Completed Responses

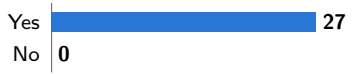
$n = 29$	n	%
Region (Q 3)		
Casablanca	8	28
Morocco (unspec.)	2	7
Oriental region	2	7
other regions (1 each)	9	31
no answer	8	28
Other daily languages (Q 4)		
English	27	93
French	24	83
Darija	20	69
Modern Standard Arabic	6	21
Tachelhit / Tamazight	5	17
Spanish	4	14

$n = 29$	n	%
Writes their language (Q 9)		
yes	23	79
no	4	14
Script used (Q 29)		
Latin	19	66
Arabic	4	14
Age (Q 12)		
20–29	16	55
$\leq$19	9	31
30–39	2	7
Education (Q 5)		
Licence	9	31
Bac	8	28
Master+	6	21
other / not disclosed	6	21

Regions hand-coded from free-text answers; “no answer” includes optionally hidden fields.

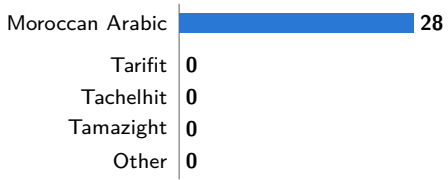
Q1–Q2: Screening

Q1 *"Is Moroccan Arabic or a Tamazight language the first language you learned to speak?"*



27 of the 29 completed responses answered.
Answering "No" together with "Other" on Q2 ended the survey.

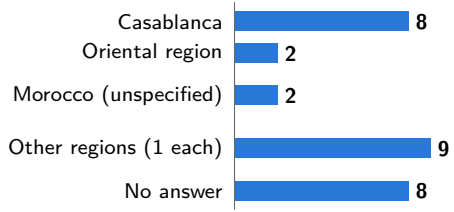
Q2 *"Which language is it? If you speak more than one of these languages, choose the one you feel closest to."*



28 of the 29 completed responses answered.

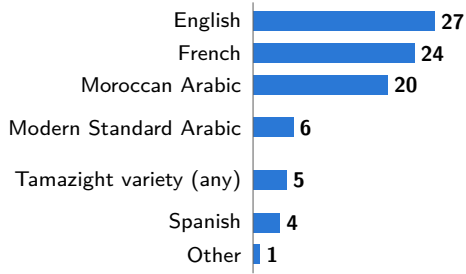
Q3–Q4: Where and Which Languages

Q3 *"In which province do you speak your language? For example, this could be your hometown."*



Free text, hand-coded. Other regions: Safi, Marrakech-Safi, Fès-Meknès, Settat/Beni Mellal, Agadir, Ouarzazate, northern Morocco, Casablanca/Souss, "my city".

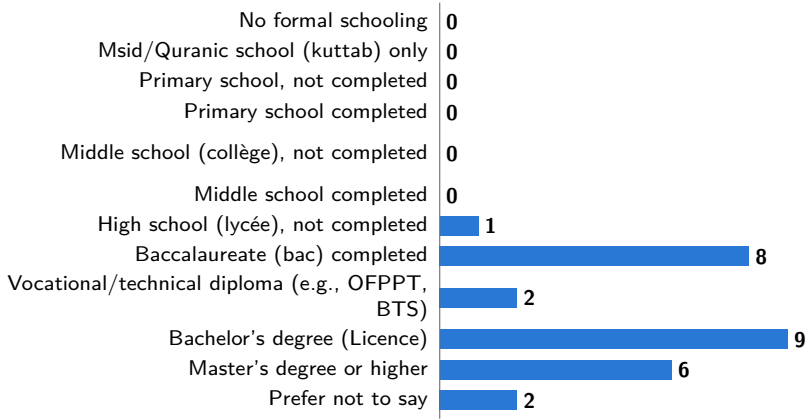
Q4 *"Which other languages do you use every day? For example, at school, at work, or with your family."*



28 of the 29 completed responses answered. Multiple answers. Tamazight options merged. "Other": Dutch.

Q5: Education

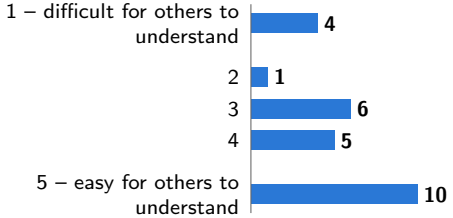
Q5 *"What is the highest level of education you completed?"*



28 of the 29 completed responses answered.

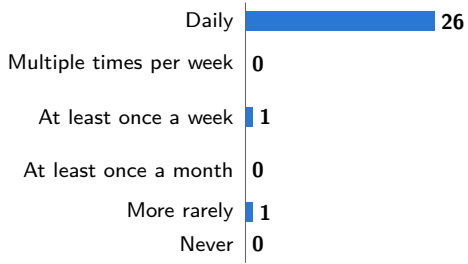
Q6–Q7: Distance and Frequency

Q6 *"How different is your language from other languages in Morocco?"*



26 of the 29 completed responses answered.

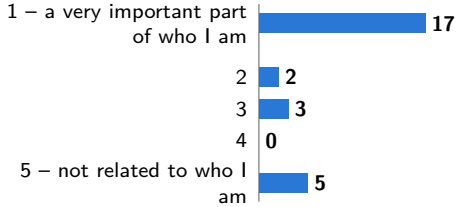
Q7 *"How often do you speak your language?"*



28 of the 29 completed responses answered.

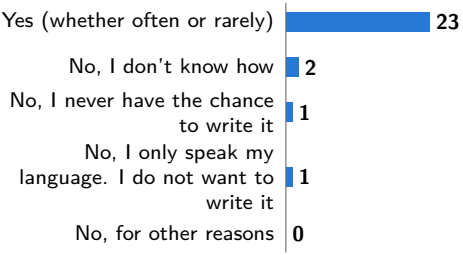
Q8–Q9: Identity and Writing

Q8 *"Do you feel that your language is an important part of who you are?"*



27 of the 29 completed responses answered.

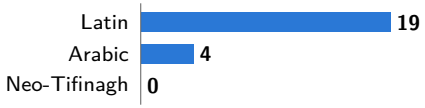
Q9 *"Do you ever write in your language?"*



27 of the 29 completed responses answered.

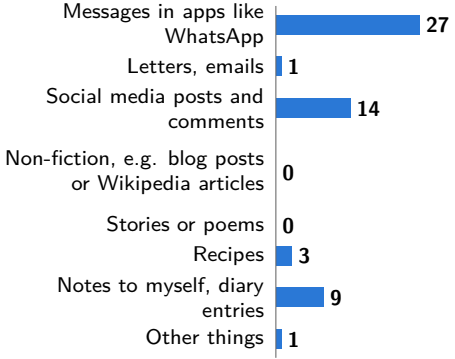
Q29 and Q10: What Respondents Write

Q29 *"Which alphabet do you use?"*



Shown only to the 23 who answered "Yes" on Q9.

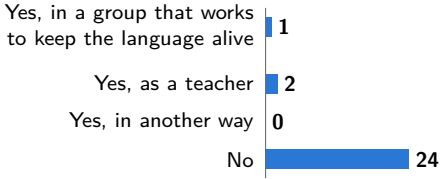
Q10 *"What do you write in your language?"*



27 of the 29 completed responses answered. Multiple answers.

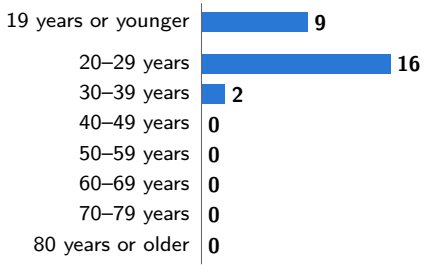
Q11–Q12: Language Work and Age

Q11 *"Do you help spread or teach your language?"*



27 of the 29 completed responses answered. Multiple answers.

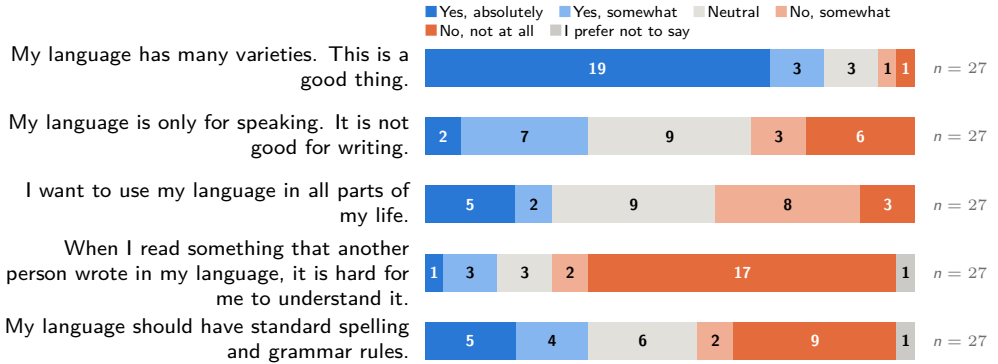
Q12 *"How old are you?"*



27 of the 29 completed responses answered.

Q14: Attitudes Toward One's Language

Q14 *"Do you agree with the following statements?"*

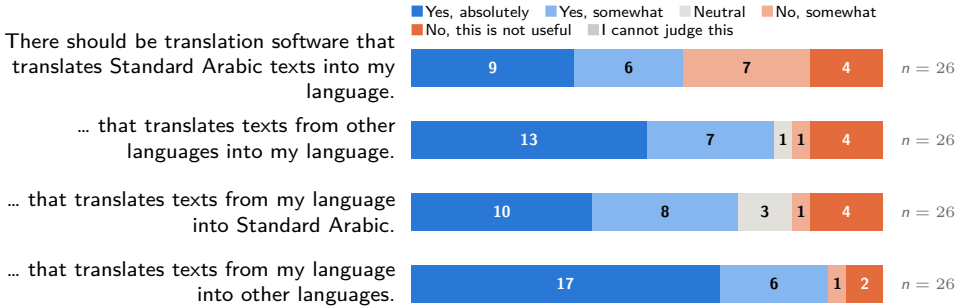


Number of respondents per answer. 27 of the 29 completed responses answered.

Q16: Translation Software

Q16 *"Do you agree with the following statements?"*

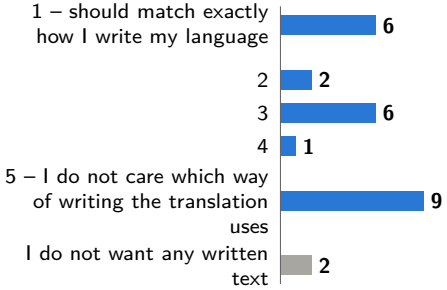
Note shown to respondents: Translation software changes text from one language into another language automatically. Two examples are DeepL and Google Translate.



Number of respondents per answer. 26 of the 29 completed responses answered.

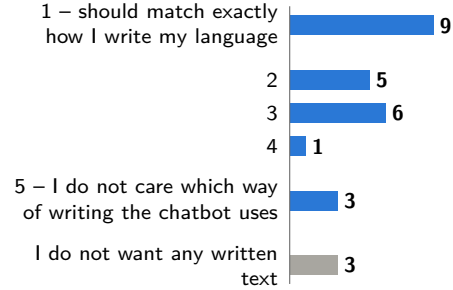
Q17 and Q22: Whose Spelling?

Q17 *"Which statement do you agree with more? Think about text that someone translates into your language."*



26 of the 29 completed responses answered.

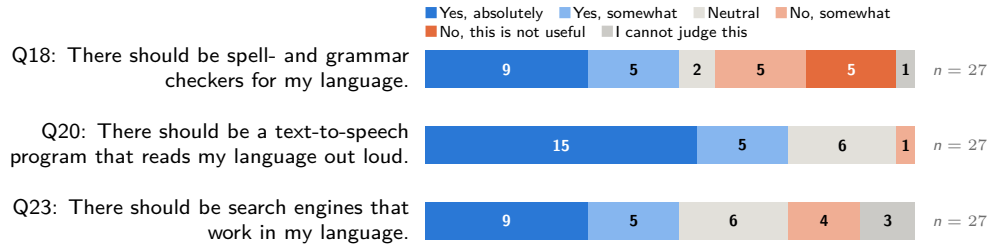
Q22 *"Which statement do you agree with more? Think about when a chatbot answers you in your language."*



27 of the 29 completed responses answered.

Q18, Q20, Q23: Spelling, Reading Aloud, Search

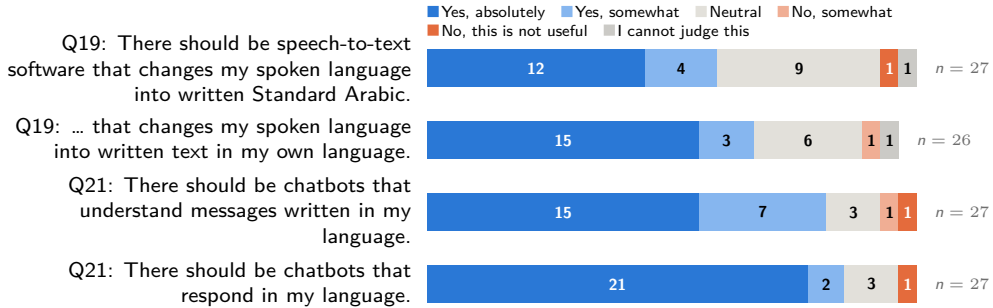
Q18, Q20, Q23 *"Do you agree with the following statement?"*



Number of respondents per answer. Each question was asked on its own page, with a short explanation and examples (Microsoft Word and WhatsApp; Google).

Q19 and Q21: Speech-to-Text and Chatbots

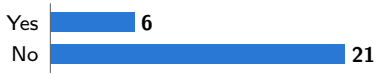
Q19, Q21 *"Do you agree with the following statements?"*



Number of respondents per answer. The chatbot question gave ChatGPT as an example.

Q25: Awareness of Language Technology

Q25 *"Do you already know about any language technology for your language?"*



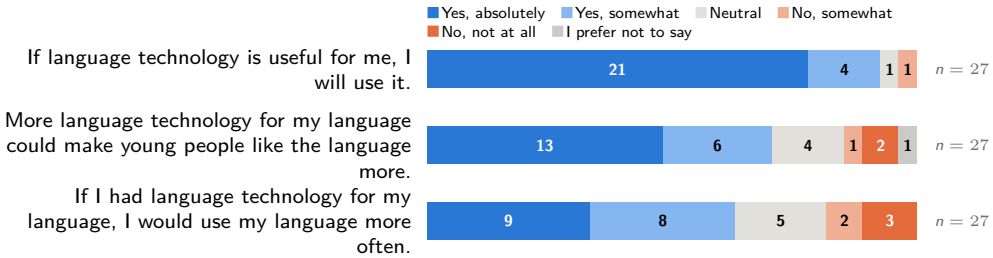
27 of the 29 completed responses answered.

Named by the 6 who said yes (Q35)

- Moulsot ASR, Gemini TTS
- Atlas-Chat, Terjman, Moulsot
- ChatGPT voice
- *kandn chatGPT kay3rf yktb b darija* ("I think ChatGPT knows how to write in Darija")
- Peace Corps Darija textbook
- learnmoroccan.com translator

Q26: Language Technology in General

Q26 *"Do you agree with the following statements?"*



Number of respondents per answer. 27 of the 29 completed responses answered.

Q28 and Open Comments

Q28 *"How did you learn about this study?"*



26 of the 29 completed responses answered. "Other": Discord (2).

Optional comments (Q13, Q15, Q24, Q27)

Answers: Q13 6, Q15 6, Q24 5, Q27 4 (several are just "No").

"autocorrect in darija would be nice"

"There can't be a rule book for Darija, because it's mostly made by and for people." (transl.)

"Darija is not suited in any way for work."

"Most AI apps like ChatGPT or Claude already do this job." (transl.)

"Well done, guys, great work." (transl.)