

## Lexical Doublets in Mixed Languages: Diagnostic Evidence from Seven Contact Varieties

Ekaterina M. Gridneva  
HSE University (Moscow)  
[egridneva@hse.ru](mailto:egridneva@hse.ru)

The current paper examines lexical doublets — synonymous pairs derived from different lexifier languages — in seven established mixed languages. It addresses four research questions: whether doublet frequency can serve as a quantitative measure of lexical mixture; what the semantic distribution of doublets (measured against WOLD borrowed scores) reveals about their status; to what extent doublet frequency correlates with the degree of grammatical mixing; and how the social functions of doublets vary across different patterns of language emergence. Table 1 below provides examples of doublets found in the mixed languages discussed in this study.

Table 1: Doublets in Mixed Languages

<b>Semantic Field (WOLD)</b>	<b>Meaning</b>	<b>French</b>	<b>Cree</b>	<b>Michif</b>
Quantity	‘all’	<i>tout</i>	<i>kahkiyaw</i>	<i>kuhkiyaw / tout</i>
<b>Semantic Field (WOLD)</b>	<b>Meaning</b>	<b>Kriol</b>	<b>Gurindji</b>	<b>Gurindji Kriol</b>
The body	‘foot’	<i>fut</i>	<i>jamana</i>	<i>fut / jamana</i>
Kinship	‘woman’	<i>woman</i>	<i>kirri</i>	<i>kirri / woman</i>
<b>Semantic Field (WOLD)</b>	<b>Meaning</b>	<b>Local Quechua</b>	<b>Spanish</b>	<b>Media Lengua</b>
Animals	‘dog’	<i>ashku</i>	<i>perro</i>	<i>ashku / pirru</i>

<b>Semantic Field (WOLD)</b>	<b>Meaning</b>	<b>Russian</b>	<b>Aleut</b>	<b>Medny Aleut</b>
Quantity	‘one’	<i>один</i>	<i>атаҕан</i>	<i>odin / ataqan</i>

The study draws on the Swadesh-100 list as a uniform diagnostic tool for seven mixed languages: Angloromani, Media Lengua, Medny Aleut, Old Helsinki Slang, Michif, Light Warlpiri, and Gurindji Kriol. The Swadesh list, despite its limitations in part-of-speech distribution — a relevant factor for Michif, for instance — provides a cross-linguistically comparable lexical sample. Moreover, its use does not imply the application of comparative-historical methods, which yield misleading results when applied to contact languages. Lexical information was gathered from specialised dictionaries, descriptive studies, and previously published Swadesh lists. The investigation proceeds along four analytical dimensions: (i) establishing doublet frequencies; (ii) assessing their semantic distribution via WOLD Borrowed Scores; (iii) examining correlations with grammatical mixing typologies; (iv) interpreting sociolinguistic functions.

The results fall into the core observations. First, the proportion of doublets is consistently predicted by how the language came into being, not by how long ago it emerged — a finding at odds with Auer’s (1999) diachronic model. The highest figures belong to Gurindji Kriol and Light Warlpiri, both developed from code-switching. Michif, a product of language intertwining, shows a moderate level. The lowest range characterises languages shaped by relexification or its mirror-opposite variant<sup>1</sup> — Media Lengua, Angloromani, and Medny Aleut. Old Helsinki Slang falls between the latter two, reflecting its interpretation as paralexification. Second, doublets do not distribute evenly: they cluster precisely in those semantic domains known to resist lexical borrowing, such as body parts and kinship terminology (WOLD scores 0.17–0.23). This inverse relation to general borrowability suggests that doublets are not fleeting borrowings but entrenched elements of the core lexicon. Third, the incidence of doublets tracks the depth of structural mixing. Languages with substantial grammatical input from both sources display high doublet counts; those with a sharp division

---

<sup>1</sup> I use the term ‘mirror-opposite relexification’ following Croft [2003] and Chlenov [2021] to describe the pattern observed in Angloromani and Medny Aleut, where — in contrast to classical relexification (lexical replacement with grammatical retention) — the lexicon of the ancestral language is preserved. I have discussed this terminological choice in more detail in a previous study [Gridneva: forthcoming], arguing that it captures the systematic relationship between these languages and the better-known relexification type.

between grammar and lexicon show few. This pattern extends to the pragmatic level (iv): in code-switching-derived varieties doublets function as active sociolinguistic resources, while in relexification-derived systems they remain largely inert.

The findings indicate that lexical doublets are not marginal or irregular phenomena, but rather systematic features of mixed languages. Their frequency and distribution are shaped by the specific historical and sociolinguistic conditions under which a given language emerged, and remain stable once the variety has been conventionalised. This does not support Auer's (1999) hypothesis that lexical variation necessarily diminishes over time. The analysis suggests that doublet patterns can serve as a diagnostic tool for the typological classification of mixed languages.

## References

- Auer 1999 — P. Auer. From code switching via language mixing to fused lects: toward a dynamic typology of bilingual speech. *International journal of bilingualism*. 1999. Vol. 3. Iss. 4. P. 309–332.
- Bakker 1997 — P. Bakker. *A language of our own: The genesis of Michif, the mixed Cree-French language of the Canadian Métis*. New York: Oxford University Press, 1997.
- Chlenov 2021 — M. A. Chlenov. A 150-Year-Old Kuril Islands Tragedy: Yet Another Solution to the Copper Island Aleut Enigma. *Études Inuit Studies*. 2021. Vol. 45. Iss. 1–2. P. 207–231.
- Croft 2003 — W. Croft. Mixed languages and acts of identity: An evolutionary approach. In Y. Matras, P. Bakker (eds.). *The Mixed Language Debate: Theoretical and Empirical Advances*. Berlin: Mouton de Gruyter, 2003. P. 41–72.
- Golovko 1994 — E. V. Golovko. Mednij Aleut or Copper Island Aleut: An Aleut-Russian mixed language. In P. Bakker, M. Mous (eds.). *Mixed Languages: 15 case studies in language intertwining*. Amsterdam: IFOTT, 1994. P. 113–121.
- Grant 1998 — A. Grant. Romani words in non-standard British English and in English slang. *The Romani element in non-standard speech*. Wiesbaden: Harrassowitz, 1998. P. 165–191.
- Gridneva 2022 — E. M. Gridneva. Mixed languages: Patterns of lexification of the core vocabulary. *Acta Linguistica Petropolitana*. 2022. Vol. 18. Iss. 3. P. 50–71.
- Gridneva (forthcoming) — E. M. Gridneva. The Hypothesis of the Dependence of the Lexical Nature of Mixed Languages on the Patterns of Their Emergence. *Vestnik Tomskogo Gosudarstvennogo Universiteta. Filologiya*. (Accepted for publication on December 4, 2025; to be published in the second half of 2026.).
- Haspelmath, M., Tadmor, U. (eds.) 2009 — M. Haspelmath, U. Tadmor (eds.). *The World Loanword Database (WOLD)*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available online at <https://wold.clld.org>.
- Jarva 2008 — V. Jarva. Old Helsinki Slang: Language mixing as a prerequisite for the formation of a new urban dialect. In M. Palander, H. Riionheimo, V. Koivisto (eds.). *Towards a New Grammar of Finnish*. Joensuu: University of Joensuu, 2008. P. 53–68.

Matras 2002 — Y. Matras. *Romani: A Linguistic Introduction*. Cambridge: Cambridge University Press, 2002.

Meakins 2008a — F. Meakins. Mixed Languages. In: *The Handbook of Language Contact*. S. 1.: Wiley-Blackwell, 2008.

Meakins 2008b — F. Meakins. *Case-marking in contact: The development and function of case morphology in Gurindji Kriol, an Australian mixed language*. Doctoral dissertation. Melbourne: University of Melbourne, 2008.

Meakins 2018 — F. Meakins. Mixed languages. *Oxford Research Encyclopedia of Linguistics*. Oxford: Oxford University Press, 2018.

Meakins, O'Shannessy 2005 — F. Meakins, C. O'Shannessy. Possessing variation: age and inalienability related variables in the possessive constructions of two Australian mixed languages. *Monash University Linguistics Papers*. 2005. Vol. 4. Iss. 2. P. 43–63.

Muysken 1981 — P. Muysken. Halfway between Quechua and Spanish: The case for relexification. In A. Highfield, A. Valdman (eds.). *Historicity and variation in creole studies*. University of Amsterdam, 1981. P. 52–78.

O'Shannessy 2005 — C. O'Shannessy. Light Warlpiri: A new language. *Australian journal of linguistics*. 2005. Vol. 25. Iss. 1. P. 31–57.

Swadesh 1971 — M. Swadesh. *The Origin and Diversification of Language*. Chicago: Aldine Atherton, 1971.