

- Dostosowanie \LaTeX a składu w innym języku
- Narzędzia polonizacji \LaTeX a
- Wzorce dzielenia wyrazów
- Czy Babel jest legalny?
- Wniosek (smutny)
- Wzorce dzielenia wyrazów
- Zmiana języka we fragmencie dokumentu
- Polskie kroje pisma
- Sposoby wprowadzania znaków narodowych
- Standardowe polecenia akcentowe \LaTeX 2 ϵ
- Tabele konwersji kodów wejściowych
- Konwersja przez znaki aktywne
- Notacja prefiksowa
- Polskie zwyczaje typograficzne
- Matematyka po polsku
- Podsumowanie
- Przykład
- Zadania \LaTeX a
- Gdzie kończą się możliwości \LaTeX a

Marcin Woliński

L^AT_EX 2_ε a sprawa polska



- zasady przenoszenia wyrazów
- dostęp do znaków narodowych w fontach
- wprowadzanie znaków narodowych
- tłumaczenie napisów wstawianych automatycznie przez \LaTeX a
- odmienność zwyczajów typograficznych

PL_AT_EX — zbiór narzędzi do składu dokumentów w języku polskim (Mariusz Olko, Marcin Woliński)

Babel — pakiet do składu dokumentów wielojęzycznych ze wsparciem m.in. dla języka polskiego (Johannes Braams)

Składniki L^AT_EXa:

- pakiet polonizacyjny `po1ski.sty`,
- pliki definicyjne dla fontów PL (`*.fd`),
- definicje dodatkowych „stron kodowych” pakietu `inputenc` (`mazovia` i `amigap1`),
- pliki konfigurujące format (`hyphen.cfg`, `language.dat` i `plhyph.tex`),
- pakiet `plprefix` obsługujący uogólnioną notację prefiksową,
- pakiet `ot1patch` pozwalający na sztukowanie polskich znaków w przypadku braku polskich fontów.

P_ET_EX:

```
\usepackage{poliski}
```

Babel:

```
\usepackage[polish]{babel}
```

lub

```
\usepackage{polish}
```

Języki, dla których ma być dostępne automatyczne dzielenie wyrazów przy przenoszeniu, trzeba wybrać na etapie generowania pliku formatu.

W trakcie pracy $\text{T}_{\text{E}}\text{X}$ a będą dostępne tylko te wzorce dzielenia, które przewidziano w formacie.

Współczesne dystrybucje T_EXa oferują mechanizm konfigurowania wzorców przenoszenia wyrazów wykorzystujący pliki konfiguracyjne z pakietu babel.

Lista języków i odpowiadających im nazw plików z wzorcami znajduje się w pliku `language.dat`. Jest on czytany w trakcie generowania formatu.

Przykładowy plik `language.dat`:

```
american hyphen.tex
=english
french frhyphen.tex
german dehyph.tex
polish plhyph.tex
```

W dystrybucjach opartych na teT_EXu można zlecić ponowne wygenerowanie wszystkich plików formatów poleceniem `fmtutil --all`

Twierdzenie: *Większość dystrybucji T_EXa zawiera nielegalną modyfikację L^AT_EXa 2_ε.*

Dowód:

Większość dystrybucji T_EXa zawiera preinstalowany pakiet Babel.

Licencja L^AT_EXa 2_ε zabrania dodawania do formatu nowych poleceń — zestaw poleceń Standardowego L^AT_EXa jest ściśle określony.

Babel w pliku konfiguracyjnym globalnie definiuje nowe polecenia.

Zatem format wygenerowany przy pomocy pliku konfiguracyjnego Babela nie ma prawa używać nazwy `latex.fmt`.

W większości dystrybucji tak właśnie jest, q.e.d.

Wszyscy autorzy pakietów są równi,
ale niektórzy autorzy są bardziej równi niż inni.

Format uzyskany przy pomocy standardowego mechanizmu konfiguracyjnego pozwala na poprawne przetwarzanie zarówno dokumentów wykorzystujących pakiet babel jak i dokumentów stosujących pakiet polski.

W przypadku braku wzorców dla jakiegoś języka użytego w dokumencie pakiet polski zgłasza błąd i blokuje automatyczne przenoszenie wyrazów w tym języku.

Babel w tej sytuacji wypisuje ostrzeżenie w pliku `.log` i używa angielskich (!) reguł przenoszenia.

Poprawne dzielenie przez \TeX a wyrazów przy przenoszeniu jest możliwe tylko przy stosowaniu fontów zawierających polskie znaki.

PLATEX: Makro `\selecthyphenation` przełącza dzielenie wyrazów na reguły właściwe dla danego języka. O inne elementy właściwe dla języka trzeba zadbać oddzielnie.

```
\selecthyphenation{english}
```

Babel: Wyrafinowany system poleceń i środowisk zmieniających wszystkie parametry zależne od języka. Podstawowe polecenie:

```
\selectlanguage{english}
```

W chwili obecnej w \LaTeX 2_ε można łatwo korzystać z następujących fontów zawierających dobrej jakości polskie znaki:

- fonty PL — polska adaptacja fontów Computer Modern dokonana przez Bogusława Jackowskiego i Marka Ryćko,
- fonty EC — European Computer Modern — „europejskie” rozszerzenie fontów Computer Modern,
- fonty PC — polska adaptacja fontów Computer Concrete (Jackowski, Ryćko),
- Antykwa toruńska, Antykwa Półtawskiego — repliki polskich krojów pisma (Janusz M. Nowacki, Bogusław Jackowski i Piotr Strzelczyk),
- QuasiTimes, QuasiPalatino, QuasiHelvetica, QuasiHelveticaCondensed, QuasiCourier, QuasiBookman, QuasiChancery — adaptacja krojów URW+ (B. Jackowski, J. Nowacki i P. Strzelczyk).

krój	układ	<i>\fontfamily</i>	pakiet	
Computer Modern	OT4, T1	cmr	<i>domyślne</i>	te
Computer Concrete	OT4, T1	ccr	beton	te
Antykwa Toruńska	OT4, QX	antt	antyktor	tl
Antykwa Półtańskiego	OT4, QX	antp	antpol	tl
QuasiTimes	OT4, QX	qtm	qtimes	tl
QuasiPalatino	OT4, QX	qpl	qpalatin	tl
QuasiHelvetica	OT4, QX	qhv	qhelveti	wp
QuasiCourier	OT4, QX	qcr	qcourier	wp
QuasiBookman	OT4, QX	qbk	qbookman	wp
<i>QuasiChancery</i>	OT4, QX	qzc	qzapfcha	wp

Dostępność: te — te_{TeX} , fp_{TeX} , TeXlive , mik_{TeX} ; tl — TeXlive ;
wp — w przygotowaniu (prawdopodobnie TeXlive 6).

Przełączenie \LaTeX 2_ε na układ fontów zawierający polskie znaki:

```
\usepackage[OT4]{fontenc}
```

Przy użyciu pakietu polski domyślne.

```
\usepackage[T1]{fontenc}
```

W przypadku pakietu polski wystarczy opcja pakietu:

```
\usepackage[T1]{polSKI}
```

- standardowe polecenia akcentowe,
- notacja naturalna (8-bitowa)
 - tabele konwersji na poziomie implementacji,
 - znaki aktywne
- notacja prefiksowa,

```
\k{a} \'c \k{e} \l{} \'n \'o \'s \'z \.z  
\k{A} \'C \k{E} \L{} \'N \'O \'S \'Z \.Z
```

W \LaTeX u 2_ε polecenia akcentowe mają znaczenie „użyj najlepszej metody uzyskania danego znaku dostępnej w bieżącym układzie fontu”. Jeżeli w bieżącym foncie jest dostępny znak z akcentem, zostanie on użyty bezpośrednio. Jeżeli go nie ma, zostanie użyte polecenie pierwotne `\accent` do położenia akcentu nad literą. Jeżeli w użyciu byłby układ nie zawierający któregoś z potrzebnych znaków składowych, \LaTeX zgłosiłby błąd.

Większość implementacji \TeX a zapewnia konwersję znaków wejściowych na postać wewnętrzną zanim \TeX je naprawdę zobaczy.

Zalety: Pliki tworzone przez \TeX a i komunikaty o błędach są czytelne. Przetwarzanie tekstów polskich odbywa się zauważalnie szybciej.

Wady: Zależne od implementacji. Poważne utrudnienie korzystania w jednym dokumencie z fontów w różnych układach.

T_EX wewnętrznie pracuje z tekstem zakodowanym wg. pozycji znaków w używanym foncie (na tej reprezentacji operuje np. algorytm dzielenia wyrazów przy przenoszeniu).

Dlatego tabele konwersji zależą zarówno od kodowania tekstu na wejściu jak i od układu stosowanych fontów.

W dystrybucjach opartych na web2c (te \TeX , fp \TeX , \TeX live) i mik \TeX u przekodowanie zapewnia mechanizm TCX. Jego uaktywnienie uzyskuje się przez umieszczenie w pierwszej linii pierwszego pliku czytane przez \TeX a następującego „komentarza”:

```
%& --translate-file=nazwa-tabeli-konwersji
```

Na przykład dla tabeli konwertującej ISO 8859-2 (Latin 2) na układ fontów PL (OT4):

```
%& --translate-file=il2-pl
```

Tablice konwersji dostępne w dystrybucji web2c:

strona kodowa	Układ fontów	
	OT4	T1
ISO 8859-2 (Latin 2)	i12-p1	i12-t1
CP 1250 (MS Windows)	cp1250p1	cp1250t1
CP 852 (tzw. IBM Latin 2)	cp852-p1	cp852-t1
Mazovia	maz-p1	
Amiga PL	amiga-p1	

W DOS-owej implementacji em \TeX tablicę konwersji trzeba zaszyć w pliku formatu (opcja /c ini \TeX a).

Problemy: Trzeba utrzymywać oddzielne pliki formatu dla różnych tablic konwersji. W dokumencie nie ma żadnej wskazówki dla \TeX a, jaki kod został użyty.

Standardowy pakiet `inputenc` zapewnia zamianę znaków narodowych na polecenia \LaTeX 2 ϵ .

Wady: Pliki pisane przez \TeX będą miały polskie znaki zapisane w standardowych poleceniach \LaTeX . Nieczytelne pliki `.log`.

Zalety: Wykorzystanie mechanizmu poleceń akcentowych. Lepsza przenośność.

Dostępne strony kodowe pakietu inputenc:

strona kodowa	opcja pakietu
ISO 8859-2 (Latin 2)	latin2
CP 1250 (MS Windows)	cp1250
CP 852 (tzw. IBM Latin 2)	cp852
Mazovia	mazovia
Amiga PL	amigapl

Na przykład:

```
\usepackage[latin2]{inputenc}
```

Notacja prefiksowa polega na zapisywaniu znaków diakrytycznych w postaci dwóch znaków: ustalonego znaku specjalnego (prefiksu) i litery łacińskiej podobnej graficznie do odpowiedniej polskiej.

Notacja pakietu babel:

```
"a "c "e "l "n "o "s "z "r  
"A "C "E "L "N "O "S "Z "R
```

Notacja pakietu polski:

```
/a /c /e /l /n /o /s /x /z  
/A /C /E /L /N /O /S /X /Z
```

Notację prefiksową w pakiecie polski uaktywnia się poleceniem `\prefixing`, wyłącza poleceniem `\nonprefixing`. Są to deklaracje lokalne.

W pakiecie babel notacja prefiksowa jest domyślnie włączona.

PDF_{TEX} zawiera dodatkowy pakiet `plprefix` pozwalający obsłużyć różne warianty notacji prefiksowej.

W wywołaniu pakietu należy podać jako opcję preferowany sposób zapisu litery ż. Na przykład:

```
\usepackage["r]{plprefix}
```

opcja	notacja polskich znaków								
/z	/a	/c	/e	/l	/n	/o	/s	/x	/z
/r	/a	/c	/e	/l	/n	/o	/s	/z	/r
"z	"a	"c	"e	"l	"n	"o	"s	"x	"z
"r	"a	"c	"e	"l	"n	"o	"s	"z	"r
@z	@a	@c	@e	@l	@n	@o	@s	@x	@z
@r	@a	@c	@e	@l	@n	@o	@s	@z	@r

L^AT_EX, według amerykańskich zwyczajów typograficznych, kładzie po kropce odstęp większy od zwykłego odstępu międzywyrazowego.

P_ET_EX automatycznie wyłącza ten mechanizm.
W przypadku użycia babilonu trzeba jawnie użyć polecenia `\frenchspacing`.

Polskie cudzysłowy są dostępne zarówno w układzie OT4 jak i T1 poprzez ligatury:

„polskie”	„polskie”
«francuskie»	<<francuskie>>

Według polskich reguł ortograficznych dywiz w słowach złożonych (np. biało-czerwony) w przypadku łamania na nim wiersza pojawia się na końcu wiersza i ponownie na początku nowego.

PE_TE_X: makro `\dywiz`

`biało\dywiz czerwony`

Babel: brak

	polski	\LaTeX
<code>\tg</code>	tg	
<code>\ctg</code>	ctg	
<code>\tgh</code>	tgh	
<code>\ctgh</code>	ctgh	
<code>\xgeq</code>	\geq	
<code>\xleq</code>	\leq	
<code>\tan</code>	tg	tan
<code>\tanh</code>	tgh	tanh
<code>\arcsin</code>	arc sin	arcsin
<code>\arccos</code>	arc cos	arccos
<code>\arctan</code>	arc tan	arctan
<code>\leq</code>	\leq	\leq
<code>\geq</code>	\geq	\geq

Powyższe definicje, redefinicje i załadowanie polskich fontów matematycznych uaktywnia się opcją `plmath` pakietu `polski`.

Ponieważ opisane redefinicje nie zawsze są pożądane, pakiet udostępnia blokującą je opcję `nomathsymbols`.

Oprócz wywołania pakietu polonizującego dokument po polsku musi zawierać informację

- o kodowaniu znaków na wejściu,
- o fontach, które mają być użyte.

Przykład dokumentu z użyciem strony kodowej ISO 8859-2 (Latin-2) (z tablicą konwersji) i fontów PL (układ OT4):

```
%& --translate-file=i12-pl
\documentclass{article}
\usepackage{polski}

\begin{document}
  Pójdź kińże tę chmurność w~głęb flaszY.
\end{document}
```

Przykład dokumentu z użyciem strony kodowej ISO 8859-2 (Latin-2) (pakiet inputenc) i fontów EC (układ T1):

```
\documentclass{article}
\usepackage[T1]{polski}
\usepackage[latin2]{inputenc}

\begin{document}
  Pójdź kińże tę chmurność w~głąb flaszki.
\end{document}
```

Przykład dokumentu z użyciem strony kodowej ISO 8859-2 (Latin-2) (pakiet inputenc) i fontów EC (układ T1):

```
\documentclass{article}
\usepackage[T1]{fontenc}
\usepackage[polish]{babel}
\usepackage[latin2]{inputenc}

\begin{document}
  Pójdź kińże tę chmurność w~głąb flaszki.
\end{document}
```

Przykład dokumentu w języku polskim, z użyciem notacji prefiksowej i fontów PL:

```
\documentclass{article}
\usepackage{polski}
\prefixing

\begin{document}
  Za/z/o/l/c g/e/s/l/a ja/x/n.
\end{document}
```

- dostęp do fontów zawierających polskie znaki akcentowane,
- poprawne przenoszenie słów polskich,
- obsługa zamiany minuskuł polskich liter na majuskuły i odwrotnie,
- w implementacji \TeX a z konfigurowalną tabelą kodów wejściowych umożliwienie definiowania makr o nazwach zawierających „polskie znaki”,
- obsługa notacji prefiksowej polskich znaków,
- właściwe polskie napisy generowane automatycznie,
- poprawne przenoszenie wyrazów złożonych typu biało-czerwony,
- polskie symbole funkcji trygonometrycznych $\backslash tg$, $\backslash ctg$, $\backslash tgh$, $\backslash ctgh$ i $\backslash arc$ i polskie znaki większe-równe i mniejsze-równe,
- zmiana znaczenia standardowych nazw $\backslash tan$, $\backslash arcsin$ itd. na polskie odpowiedniki.

P \LaTeX nie próbuje ingerować w projekt typograficzny dokumentu. Wymagałoby to bowiem uzależnienia pakietu od standardowych klas dokumentów L \TeX a.